

SEARCH REQUEST FORM

Requestor's Name: _____ Serial Number: _____
Date: _____ Phone: _____ Art Unit: _____

Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

STAFF USE ONLY

Date completed: 10-07-02
Searcher: BEVERLY C 4994
Terminal time: 25
Elapsed time: _____
CPU time: _____
Total time: 25
Number of Searches: _____
Number of Databases: 1

Search Site

_____ STIC
_____ CM-1
_____ Pre-S

Type of Search

_____ N.A. Sequence
_____ A.A. Sequence
_____ Structure
_____ Bibliographic

Vendors

_____ IG Suite
_____ STN
_____ Dialog
_____ APS
_____ Geninfo
_____ SDC
_____ DARC/Questel
_____ Other CGN

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 02:19:23 ; Search time 1114.8 seconds
(without alignments)
11996.956 Million cell updates/sec

Title: US-09-736-968A-1
Perfect score: 6372
Sequence: 1 gacgcgagacacatgctgc.....aagcaaaaaaaaaaaaaa 6372

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank:
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_cm:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hlg_hum:*
31: em_hlg_inv:*
32: em_hlg_other:*
33: em_hlgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result Query Score Match Length DB ID Description

1	6372	100.0	6372	6	AX173175	AX173175 Sequence
2	4882.8	76.6	4886	9	AB037816	AB037816 Homo sapi
3	2797.8	43.9	2825	9	BC008335	BC008335 Homo sapi
4	2140	33.6	2149	6	AX173233	AX173233 Sequence
5	2132.8	33.5	6828	6	AX172874	AX172874 Sequence
6	1729.4	27.1	7215	6	AX174569	AX174569 Sequence
7	1468.2	23.0	4311	6	AB051558	AB051558 Homo sapi
8	1446.6	22.7	4144	6	AX172880	AX172880 Sequence
9	1446.6	22.7	4200	6	AX172932	AX172932 Sequence
10	1446.6	22.7	4144	6	AX172980	AX172980 Sequence
11	1357.6	21.3	1900	10	BC021414	BC021414 Mus muscu
12	1225.4	19.2	3327	6	AK055905	AK055905 Homo sapi
13	1207	18.9	3253	6	AX172978	AX172978 Sequence
14	1167.8	18.3	4577	9	AK024436	AK024436 Homo sapi
15	1031.8	16.2	4026	6	AX174571	AX174571 Sequence
16	1031.8	16.2	4027	6	AX174664	AX174664 Sequence
17	1007	15.8	2452	9	AK055493	AK055493 Homo sapi
18	865.8	13.6	2014	6	AX174662	AX174662 Sequence
19	815.6	12.8	2904	9	BC019102	BC019102 Homo sapi
20	597.8	9.4	2607	9	AK055401	AK055401 Homo sapi
21	408.4	6.4	73515	2	AC017375	AC017375 Drosophi
22	407.4	6.4	166626	3	AC008318	AC008318 Drosophi
23	407.4	6.4	301639	3	AE003590	AE003590 Drosophi
24	400.8	6.3	1242	9	AF194407	AF194407 Homo sapi
25	377.2	5.1	434	6	AX329840	AX329840 Sequence
26	322.4	5.1	3899	6	AX058220	AX058220 Sequence
27	308.4	4.8	2998	10	BC009134	BC009134 Mus muscu
28	274.6	4.3	3227	10	RNTRG	AB028981 Homo sapi
29	255.2	4.0	5862	9	AB028981	AB028981 Homo sapi
30	255.2	4.0	7506	6	AX255048	AX255048 Sequence
31	254.8	4.0	123321	9	AC011472	AC011472 Homo sapi
32	249	3.9	37107	9	AC009000	AC009000 Homo sapi
33	247.8	3.9	1669	9	HSM802936	AX173254 Sequence
34	238.2	3.7	321	6	AX173254	AX173254 Sequence
35	238.2	3.7	3023	9	AK022412	AK022412 Homo sapi
36	216.6	3.4	281	6	AX173240	AX173240 Sequence
37	213.8	3.4	1605	6	AX174666	AX174666 Sequence
38	203.6	3.2	2413	9	BC015018	BC015018 Homo sapi
39	199.8	3.1	2036	9	AK001253	AK001253 Homo sapi
40	198.4	3.1	2299	9	AK054649	AK054649 Homo sapi
41	198.2	3.1	2422	9	AK000227	AK000227 Homo sapi
42	195.2	3.1	382	6	AX341959	AX341959 Sequence
43	194.8	3.1	2610	6	AX058192	AX058192 Sequence
44	194.8	3.1	3152	9	AK056684	AK056684 Homo sapi
45	194.8	3.1	3472	6	AX173120	AX173120 Sequence

ALIGNMENTS

RESULT	1					
LOCUS	AX173175	6372 bp	DNA	linear	PAT 03-JUN-2001	
DEFINITION	Sequence 1 from Patent WO0142295.					
ACCESSION	AX173175					
VERSION	AX173175.1	GI:14598036				
KEYWORDS						
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
AUTHORS	Lu, P., Garman, J.D. and Candia, A.F.					
TITLE	Clasp-7 transmembrane protein					
JOURNAL	Patent: WO 0142295-A 1 14-JUN-2001; Arbor Vita Corporation (US)					
FEATURES	Location/Qualifiers					
source	1..6372					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/note="full length human CLASP-7 cDNA"					
	13..6156					
	/note="human CLASP-7"					

1026a) H

```
/codon_start=1
/db_xref="GI:14598037"
/translation="MASSERRAFANKINRTVAEVRKQVSRERSSPHSSRRSSSLG
VPLREVEVPEDEEDLVLRPPDAEPGLRDVLVEFPDDIDELLQRECKRTTGPJPKD
EKLDQVRAAVEMVIEDWYIVHRYQYLSAAYSPVDDIDREOKRQKLPQVVEODAG
DERSPEDNSRRSGSPEDTPRSSASCIPLRLNADSLPLSLAEAEEDNR
NETLRORHPALLTLVPADEDVAERCPREPHNGORLYVCLSLKEIEIEP
IFGLIAYDVERKKIISENFPDLSNMSGILRAGCTAPALSTLARSJIEVATPSP
DIRITILEVLOOGDISSECEPYMLKEDTKAKNKEIKELRLAEQDCTLGKRM
PFAVTAVHLANIVSSAGLDROSDSBERKPATDTRKRGPDORASGGDACSFGFR
PATLVTFNEFKOEAEKLSDEDLKFLADMRPSLLRLRPVTAQLKIDISAPNPH
FCLSPDLHLIKPYDPDRGRPKELIEPAKEVYAPRTSYRNLYVPHDLNSRQGS
VRNLAVVOYMTGEDPSQALPIFGKSSCEPTEARETEVYVHNKSPFEYEFKILP
ACVTENHLLLETFYHVSQCPRGTALETPIVGFMTWILDHGRKTCFCLPVSQDP
PSYSLTPDALPDMRWGDKGVSELTGAVSVEGDPYLDKFEPTLVHVEEGAP
FRKDVTLSGNEOELRASLALRLASPRIVAFSHVLDKRVITVPTIISOIV
NLGCAFEAMAHVSYIHRSLAEAOARICPOLAYVYHARLPETPEPSLPDGPV
TVQAKTILARSGRPASLTILARSKSISSNPDILAVAGSVDDVESRLAKLHEBEL
QWVSSAVREBALIQHAWPEFQILMVKSMALHLIGRLDTPKRLRPGKFLDITLALV
GSLVEITRVHNDVLEHNLASLAEFLSDLSLDVDRGFVSLVAHKQVATRILOS
SPNPAALITLMEPTRIKLSHEHYVTLNLPCLSPAPSPSVSTSSQSTFSSOA
PDPKVTSMFELSGPFRQHFPLAGLITLLELALAEPAEGAPLHKKATSAVSLICGH
DTPDRVATVAKVAVELYLPLISTARBDPLPHDPAEGRSGRSRLASMLSDTEGSG
DIAGTINPSVAMALAGPRLA PGSRASISQSPPTASAGALSBSRTILACYVLYLK
NTEPALIQNAIDITLPLQGRLLDLHLICLAETVIGKKAERKINSLFKKSLDKAR
LEBAIILGTICARODMVRKRSRSPFGNPEVNRKRSVTHMKQISDHVDKTKMEHEA
LVEGNLITAEASLVLDLLEIIVQTVMLSEARVESLVGLAVLYSLGASQSLAFLOHG
LATGRALVSEFPELLPEDEDELICADLCILRLHRCGSRISTIRTHASASLYLIMRQNE
IGHNFAVRKQVMTSLSVGTQNSFESEHLRSRLITLYAEEDMLDSTFAEOV
DLNENLHMILLITVYKMEHQEDPEMLDILMYRIARQYSGPDLRLIWMONMKRAEL
GNHAEAOCHVAALAEYLALEALLLEDOHRLPVCSCFQWISLSEESIISDILSPD
EEGCSKHPTELGLVGLLEQANGYFTMGSLYAVAVENYKNLPILEHARDYKRIAAV
HGKIQEAFIKIMHOSGWERFEGTFRVNGYGHFEDLDEQETVKEPSTILAEISH
RLSEFYTERGDDVVELIKDSTYVDSKLSQSAVYIQTIVYEPETVYLYKRVTFD
RNYGLRTFLCTPPTPDRAHAGELPDQHKRKLTLSDHAFPIYKTRIRKCHEEVLPI
PVEVALIEDMOKTRELAFATODPDAKMIOMYLQGSVPTQKGEVLEAVQYFLAEP
EDPLFRHNNKRLKCFDKCEKEDALRKRAKILGPDQKEYHRELRNYCRLREALOP
LLIQRLPOLAMPPLPGLRNSLNASFPKADL"
BASE COUNT      1273 a      2050 c      1849 g      1200 t
ORIGIN

Query Match      100.0%; Score 6372; DB 6; Length 6372;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 6372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gacgcgagagaccatgctgcgtccgcgagcgccgttcgcgcacaaagatcaacaagag 60
        |||||||
DB      1 GACGCGAGGACCATGGCTCTCGAGCGCGCCCTTCGGCGCAAGATCAACAGAGAG 60

QY      61 gtgcgcgacaggtgcgaaagcaggtgtcccggaacgcagtgctcccccactccagc 120
        |||||||
DB      61 GTGGCGCGAGAGCTGCGGAACAGAGTGTCTCCGGGAGACGAGTGGCTCCCCCACTCCAGC 120

QY      121 aagcgctgcagcagctccctctgggggttcccaactgactgaagtgttcgagcccttggacctt 180
        |||||||
DB      121 AAGCGCTGCGAGAGCTCCCTCGGGGTCCCACTGACTGAAGTGTTCGAGCCCTCGAGACTT 180

QY      181 gaggaagtactctgagccgcccacacagatgtcagcccgggcccccacagggaccttgat 240
        |||||||
DB      181 GAGGAGTACTCTCTGAGCGCGCGCACACAGATCTGAGCCCGGCCCTCAAGGACCTGTGTA 240

QY      241 gaattcccaagctgactgtagagctgtctgtagcccccggagaaatgcgagacacagag 300
        |||||||
DB      241 GAATTCCCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY      301 cccggagatcccccaagaaactgatacccaaggttagggcgcggtggagatgatat 360
        |||||||
DB      301 CCCGGGATCCCCAAGAAACTGATGCCCAAGGTAGGGGCGCGGTGGAGATGTATA 360

QY      361 attgaggaactggatctgtccacagaaggtatcagtaactgagtgacatatcacagccc 420
        |||||||
```

1501 aagatcgacattctctcggctctctgaaaaatcccactctgcctctccctcgaagtgcct 1560
1501 AAGATCGACATTCTTCGGCTCTGTAAAAATCCCACTTCTGCTCTCCCTGAGCTGCTT 1560
1561 catatcaagccctaccggagcccccaggccggcccccagaagagatctcgyagttcccc 1620
1561 CATATCAAGCCCTACCCGAGACCCCAAGGGCCCGCCACCAAGAAATCTGTAGATTCCTCC 1620
1621 gcccggaagtctatgccccccataccagctacaggaactgctgtatagttaccggac 1680
1621 GCCCGGGAAGTCTATGCCCCCATATACAGCTACAGGAACCTGCTTACGTGTACCCGGAC 1680
1681 agcctcaactctcagcagcccgcaaggctctcgtgcgcaacctgctgctgagtgcaatc 1740
1681 AGCCTCAACTCTCAGCAGCGCGCAGGGCTCCGTCGCAACCTGTGTGAGTGCAGTAC 1740
1741 atgacagagcgagagaccccgagagctctgcgcgtacatctttggcaaatccaagctgagat 1800
1741 ATGACAGGCGAGAGACCCCAAGGCTCTGCGGTATCTTGGCAAGTCAAGCTGCAGT 1800
1801 gaattaccgcgagagcctctcacaacgctgctctacatacaagctcccccgaagtctac 1860
1801 GAATTATACCGGGAAGGCTCTTACACCGGTGCTTACCATTAACAAGTCCCGGAGTTCTAC 1860
1861 gaagagttcaagctgcatcttccaagctgctggaacagaaacatacactgctgttcaac 1920
1861 GAGAGATTCAAGCTGCATCTTCCAGCTTCGCTGACAGAAACATACCTGCTGTTCAAC 1920
1921 ttctaccatgctagctgctcagagcccccggcgagcactgccccttgagaaacacgctggagctt 1980
1921 TTCTACCATGCTACGCTGCGACGCCCGCGGCGGACTGCTTGAGAACACCCCTGGGGCTTT 1980
1981 acttgatcccaactgctgctcagcagcggcgctgagagaccgcccctctgctcccaatg 2040
1981 ACTTGATCCCACTGCTGAGAGACGAGGCGCTGAGAGACCGGCCCTTCTGCTCCACAGTG 2040
2041 tctgtgagcagcgccgcgcgcagctatctccgtgtcacaacccgagatgtgagcttccggag 2100
2041 TCTGTGAGCAGCGCGCGGCCCAAGCTATTCGCTGCTCACACCCGATGTGGCTTCGGGGC 2100
2101 atcgctgtggtgagagctgtaacaaggcgctgttcaggtgagagctcaacgctgttccctc 2160
2101 ATCGCTGTGGGTGAGAGCTGTACAAAGGCGTGTTCAGTGTGAGAGCTCACAGCCGTGTCTCT 2160
2161 gtgcaccccccaagjacccttactctgtaacaaatcttcaacctggtgcacgtlccctggagag 2220
2161 GTGCACCCCCCAGAGACCCCTAACCTGGAATAATCTTCACTGAGTGACAGTCCCTGGAGAG 2220
2221 ggaagccttcccaatcccggtctcaagaaactgtgtcgtgagcgagggcagctgtagcagag 2280
2221 GGAGCCTTCCCAATTCGGGCTCAAGGACACTGTGCTAGCGAGGGCAACGTTGAGGACGAG 2280
2281 ctgcggggcagctctgcaagaactgcctggcgagcagcccgaaaccccttctgacctctcc 2340
2281 CTGCGGGGCCAGTCTTGCAGACACTGCGCTGCGCAGCGCCGGAACCCCTTGTGGCTCTCTCC 2340
2341 caccacgtgctgtaacaagctcgtgcgtcgtgtaatacaaggcccccgatacaatgagtcag 2400
2341 CACCACGTCGTGACAAAGCTCGTGGCTGTATCATCAGGCCCGCCGATCATATGAGTGCAG 2400
2401 attgtgaacctggcgctgtagagcctttgaaagcaatgagcccaatgtaataagccttgttcaac 2460
2401 ATTGTGAACCTGGCGCTGTGAGGCTTTGAAGCAATGGCCCAATGTATGTAGCCCTGTTTCAAC 2460
2461 cggagcctgtagagcaagcccgagatgcgccggtgctactgcccacagctgctgctcgtc 2520
2461 CGGAGCCTGAGAGCAGCCAGATGCGCGGCTGACTGCCCAAGAGCTGGCTGCTTACGTC 2520
2521 cactaagccttctgccttctgtaactgagccagcctcccgagatgaggggccctccagtg 2580
2521 CACTAAGCCTTCTGCCTTCTGTGACACTGAGCCAGCCTCCGGATGGGGCCCTCCAGTGG 2580

2581 acagtgagagctgcacaaactggcccgctggtctgtgtcgcgcccgcaagcctctaccctggcg 2640
2581 ACAATGAGAGCTGCGCAACATGGGCCGTGTGTGTGGCTCCCAAGCCTTACATCTGGCG 2640
2641 cgttccaaagagcatcaagcagcaagaacccagacctgcgcgtgagccctgctgctgtgat 2700
2641 CGTTCCAAAGACATACGACAGCAACCCCTGACTGCGCTGGCCCTTGCTGTGTGAGAT 2700
2701 gacgaggttcccgcatcctcgcgcagcaagctgctctcaagaggagctgctctgcagtg 2760
2701 GACGAGGTTTCCCGCATCTGCGCAGCAGCAAGCTCTTACAGAGAGAGTGGCTGTGAGTGG 2760
2761 gtgtgtcagcagagtgctgcgtgaagcgagagccatctccagagcccgctggttcttcttcag 2820
2761 GTGTGTCAAGCAAGATGCGGTACGCGAGGCGCATCTCCAGACGCCGTGGTCTTCTTCCAG 2820
2821 ctcatgtgtgaagatgagcgctgcaactgctgctcttgccagagactagacacaccgcgc 2880
2821 CTCAATGTTGAAGATATGGCGCTGCACACTGCTGCTTGGCCAGGACTAGACACCCCGC 2880
2881 aagctgcgcttcccggaagcgttccctggaagcaatacagctgtggttggctctgtgggc 2940
2881 AAGCTGCGCTTCCCGGACGCTTCTGAGACATCACTGCTTGGTGGCTGTGTGGC 2940
2941 ctggaagttcaatcaccgctgttccaaagagatgtgagctgagcagacccctcaagcgcaagc 3000
2941 CTGGAAGTCAATCAACCGGTGTCACAAAGATGTGAGCTGGCCAGACCTCAACGCCAGC 3000
3001 ctggtcttcttctcctcagtgaaactctgtlccctggttgagccggagcttgttctcaagctg 3060
3001 CTGGCTTCTTCTTCTCAGTGAACCTTCTGCTTCCGTGTGAGCCGGGCTTGTCTTACGCTTG 3060
3061 gtccgggcccacatacaagcaggtgagccagcggtccagtcggtcccttaataccaagagcc 3120
3061 GTCCGGGCCCACTACAAAGCAGGTGGCCACGGGCTCCAGTGTCTCCTTAATCAAGCAGCC 3120
3121 ctgtgacacctgcgcataagaaattcaaccgcacatcctgtgcagccagcaactaagctgac 3180
3121 CTGTGACACCTGCGCATAGAAATTCAACCGCATCTGTGTGACGCCAGCACTACGTATGCC 3180
3181 ctcaaccttccctctgtgccccctgtlcaactctcagcactcgcctccctctgtgtctcc 3240
3181 CTCAACCTTCCCTGTGTGCCCTCTGTACCTTCACCTGACCTGCGCTTCCCTCTGTTCTCTCC 3240
3241 accaactcccaagagcttcaacttctcaagccaaagcccccgaaagggagacagatg 3300
3241 AOCACCTCCAGAGCTCCACCTTCTCCAGCCAAAGCCCCGAGCCCAAGAGTGAACGATG 3300
3301 ttggaactgaatgtaacaaattccgcgcagagcaacttccatagcttggctcctgtgaagag 3360
3301 TTGCAACTGAATGTAACAAATTCGCGCAGAGCACTTCTAGCTGGGCTTCCTGTACGGAG 3360
3361 ctgcgactgcgacctcgaaactcgagagctgtaaggggacatcctgttgcacaagaagcgatc 3420
3361 CTGCGACTGCGCCCTCGAACCTGAGCGTGAAGGGCCATTCCTGTGTGCAAGAAAGGCCATC 3420
3421 agtgtgtgcaacagctgtctatgttggccaatgacatgacatgcccccgctacgcgcgaagcgact 3480
3421 AGTGTGTGCAACAGCTCTATGTGGCCATGACACTGACCCCGCTACGCGGAAGCCACT 3480
3481 gtgaagctcgtgtggtgcgagctgtaacctgcaactgcttccgalttgacagggataacttg 3540
3481 GTGAAGGCTCTGTGTGGCGGAGCTGTACTGACACTGTTTGTGATTGCAAGGGATACCTTG 3540
3541 ccaagcgtcgaatgacttgcgtgaaggcccaagtcagcgggtcaagaactggtcctaagtctc 3600
3541 CCACGGCTGCAATGACTTGTGTGAGGGCCCAAGTCAAGCGGTCAAGATGGCTTCATATCTT 3600
3601 gactaagacaaagaaaggcgaaggagacattgcgggtaccatcaaacccctctgttggcagtg 3660
3601 GACTAAGACAAAGAAAGGCGAAGGGGACATGTGGGTACCAATCAACCCCTCTGTGGCCATG 3660
3661 gccattgctgtgtgcgccctagccccctgtctccggcgcaagcatctccaggggcacaca 3720

```
Db 3661 GCCATTGCTGGTGGCCCTTAGCCCTGCTCCGGGGCCAGCATCTCCAGGGGCCACCA 3720
Qy 3721 acggtcttcgacgaggtctgtccctctctgtctagtaaaacggagacttggtcggtc 3780
Db 3721 ACGGCTTTCGGGAGGCGGTGTGCCCTCTCTCTAGTCAAGTCAAGCCGGACTTGTGCTGT 3780
Qy 3781 gtgtctgtgtgtctgtgaaaaaacacacgagccggtcctctgcagcgctggccactgacctg 3840
Db 3781 GTGTGTGTGTGTGTGAAAAAACACCGAGCGCGGCTCCTGCGAGCGCTGGCCACTGAGCTG 3840
Qy 3841 aacactcccccaagctggagacftctgttgactgtgttaactttgctcgtgctgaccttgag 3900
Db 3841 ACACCTCCCCCAGCTGGAGCGTGTGTGACTTGTGTGACTTGTGTGACTTGTGTGACTTGTGAG 3900
Qy 3901 tacaaggggaaaaaaggccttgaacgcatacaacagctccatccataaaatctcttgat 3960
Db 3901 TACAAGGGGAAAAAGGCGCTTTGAAGCGCATACAGCCTTCATTTCAAAAAATCTTGAT 3960
Qy 3961 atgaagcgcgcgctgaagaaagcaatctgtgttaacaatcggaagctcgacaagaatggt 4020
Db 3961 ATGAAGCGCGCGCTGAAGAAAGCCATTCTGGTATCACTATCGAGCTGACAAAGAAATGTG 4020
Qy 4021 cggcgaaagtctgtgaagaaagccgcttgggaatccagagaagactgtcgctggcggaagac 4080
Db 4021 CGGCGAAAGTCTGTGAGAGAGACCCGTTTGGGAATCCAGAGAACTGGCGTGGCGGAAGAC 4080
Qy 4081 gtcaacaacttgaagcaaacctcaagaccgctgtgacaagaacaaaggaatgaatggaacac 4140
Db 4081 GTCAACAACCTGGAAGCAAAACCTCAGACCGCGTGAGAACAGCAAGATGTAATGGAACAC 4140
Qy 4141 gaagccttgtgtgaagaaagaaacctcgcaaacggaagcaagctagtgcttgcagacactg 4200
Db 4141 GAGGCTTGTGTGAAAGGAACCTGGCAACCGAGCAAGCCTAGTGGTTCGACACACTG 4200
Qy 4201 gagaatactgtgagaagcgtgactgtcttaagaagcccggaagagcgtcttgaggagagtg 4260
Db 4201 GAGATCATGCTGCAAGAGCGTGTTCGAAGCCCGGAGAGCGCTTGTGGGGGAGTg 4260
Qy 4261 ctgaaggtctgtctgttaacagcctgggcagtgcccaagctgacctctctctgcagactgac 4320
Db 4261 CTGAAGGTTGTGCTGTACACCTGGGCGAGTGCACAGAGTGCCTTCTTTCGACGATGAC 4320
Qy 4321 ctggccaccccaaggggcccttgtgtccaaagttcccggaactgtctgttcgaagaaagacag 4380
Db 4321 CTGGCCACCCCAAGGGGCCCTGTGTCTCAAGTTCGGAACCTGTGTTCAGAGAGGACAG 4380
Qy 4381 gaagctgtgtgtcgaacctgtgtcgaagctcctcaagacactgtgtgcagccgcaatcaagacc 4440
Db 4381 GAGCTGTGTGCCAGCTGTGGCCGTGAGGCTCTTACGACACTGTGTGCGACGCCGATCAAGACC 4440
Qy 4441 atccgcagcagcagcagcgccctcgctgttaactgtctcaatgaagcaagacttcgagatcg 4500
Db 4441 ATCCGACGACGACCGCCGCTCGCTGTACTGTCTCATGCGACACAAACTTCAGATTCGGC 4500
Qy 4501 caacaacttggccgtgtgaaagtgcaggtcaccaatgtctctgtcccttggctggagagc 4560
Db 4501 CAACAATTTGGCCGCTGTGAAGATGCAAGTCAACATGTCTCTGTCTGCTGGTGGGAGC 4560
Qy 4561 acgcagaactcaatgaagaagcactgtgcagcttcaactaaaacatccatactatgtc 4620
Db 4561 ACGCAGAACTTCAATGTAAGAGACCTGCGAGCTTCACCTCAAAACATCTCACTATGCT 4620
Qy 4621 gaggaggaatgtgggtctgtggagaaacagcacttcgcagagcaagctccaggaactgtctc 4680
Db 4621 GAGGAGGAGATGGGTGTGGGAGACACTTCGACAGACAGGTCCAGAGACTGATGTTTC 4680
Qy 4681 aaactgacatatcctgtgcagacaggttgaagaatgaagaagaaacaaagaagaaaccttgag 4740
Db 4681 AACCTGACATATCTCTGACGACACCGTGAAGATGAAGAGAACCAAGAGAACCTTGAG 4740
Qy 4741 atgtcatcagacctatctataagaattgcccgggtctacagaggtctacccgaccttcgg 4800
Db 4741 ATGTCTATCAGACCTCATGTAAGAAATTGCCCGGGCTACAGGCGCTACCCGACCTTGCG 4800
Qy 4801 ctgaacctgtgtcagaacaatlgccgggaagcagcggagcttgggcaacacacgcgcgaagcc 4860
Db 4801 CTGACCTGTGTGAGAAATATGGCCGGGAAGCACAGCGGAGACTGGGAAACACACCGCAGGCC 4860
Qy 4861 gcccaatgtatgttgcagccggccgacctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4920
Db 4861 GCCCAATGTATGTGTCACCGCCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4920
Qy 4921 cagcgcaaccttgcggcttgggtctgtcttccctccagaacatctcaatccacgtgtgaag 4980
Db 4921 CAGCGCACCTTGCCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 4980
Qy 4981 gagtcgcacatctccagacacatctctgtccgcgaagagagagcttctgtccgggaag 5040
Db 4981 GAGTCCGCATCTCCGACACATCTGTGTGCCGACGAGAGAGGCTTGTGTGTGTGTGTGTGT 5040
Qy 5041 caattcaactgaagcttgggtgtgaaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5100
Db 5041 CACTTCACTGAGACTTGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5100
Qy 5101 ggcgggtcttctacgaagcggtgaatgaaggtctcaagaacatccatccatcccttgaagcc 5160
Db 5101 GCGGGGCTCTACGAGCGGTGAATGAGGTCTACAAGAACCTCATGCCATCTCGGAAGCC 5160
Qy 5161 caacgtgaactacaagaatctgacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5220
Db 5161 CACCGGTGACTTCAAGAACTGTGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5220
Qy 5221 atcatgtcacacgaagttccggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5280
Db 5221 ATCATGTCAACAGAGTTCGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5280
Qy 5281 taagcgccccactctgtgtgacctgtgaatgaagcaagagttgtgtacaagagcaatcgatc 5340
Db 5281 TACGGCGCCCACTTTCGCTGACTGTGATGAGCAGAGTTTGTGTGTGTGTGTGTGTGTGTGTGT 5340
Qy 5341 acgaagctgtgcagagatctcaacccggtgtgaagatctcaagagagagattgtggagac 5400
Db 5341 ACGAAGCTGTGCGAGAGATTCACACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5400
Qy 5401 gaagctgtgtgaatlatcaaaagactcttaacctgtgtgaacaagltccaaagcttgaatcaaa 5460
Db 5401 GACGCTGTGTGAGATTAACAAGACTTTTACCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5460
Qy 5461 aaggtctacatccagatcaatgtgtgtgaacgcttactgtatatactcaagactcaagac 5520
Db 5461 AAGGCTTACATCCAGATCAGTATGTGTGAACCTGTATCTTGTATCTTACGAGCTCAAGGAC 5520
Qy 5521 cgggtgaacctacttgaacgcaactatggggtctgtgcacatctccgtgtctgcagcggttc 5580
Db 5521 CGGGTGAACCTACTTGTACCGCAACTATGGGCTTGTGCATTTCTGTCTCACGCCGTTTC 5580
Qy 5581 acgcgcgaatggggcgacacacggggagctgtcccgagcaacacaaagcgtlaagcgtgtctc 5640
Db 5581 ACGCGGATGGGGCGCACACGGGAGACTCCGACGACAAACAAGCGTAAAGCGTGTGCTC 5640
Qy 5641 agacgcagacacagcctccctacatcaaaagctcgatccgtgtgtgtgtgtgtgtgtgtgtgtgt 5700
Db 5641 AGCACCAACACAGCGCTTCCCTCATCAATCAAGATCGCATCGTGTGTGTGTGTGTGTGTGTGT 5700
Qy 5701 acggtgtctgaacgcagtgaggtgtggcatcgagagacatgcagagaagaaagacacggagctg 5760
Db 5701 ACGGTGTGTGACGGCAGTGGAGGTGGCATGTGAGAGATGTAGAAAGAACACGGAGACTG 5760
Qy 5761 gctttgtgcacgcgaagcaagcaacacagatgtctaagaatgtctacagaatgtgtctcaagggc 5820
Db 5761 GCTTTGTGCACCGAGCAGGACCCACAGATGTAAAGATGTCTAAGATGTGTGTGTGTGTGTGT 5820
Qy 5821 tctgttagggccacacgttgaaacaggttcccttggaggtggccaggtgtgttttaagcaag 5880
Db 5821 TCTGTAGGCCCAACCGTGAACAGGCTCCCTGTGAGGTGCGCCAGGTGTGTGTGTGTGTGTGT 5880
```

QY 5881 atcccggaagaccaccaagctctccgcgcatcacaaatgctgctgtctcaaggac 5940
|||||
Db 5881 ATCCGGAAGACCCCAAGCTCTTCGGCATCACAAATTGGCGCTCTGCTTCAAGAC 5940
QY 5941 ttctgcaagaatagtgaagatgctgcggaataaagaagccctgattggccggagac 6000
|||||
Db 5941 TTCTGCAAGAATGTGAGATGCTGCGGAAATAAGAGCCCTGATTGGGCGGACGAC 6000
QY 6001 aaggagatcaccctggaagtgagcgaactatgcccctgcggaagctctgagccc 6060
|||||
Db 6001 AAGGAGATCACCGTGAAGTGAGCGCACTACTGCGCTGCGGAGGCTCTGCAAGCCC 6060
QY 6061 ctgctatcccaagcgcctgcgcccagctgattgagcaccaccccccgcgctcagaagctcc 6120
|||||
Db 6061 CTGCTATCCCAAGCGCTGCCCGCCAGGTGATGCGACACCCACCGCGCTCAGAACTCC 6120
QY 6121 ttgacaagcagatgtccggaagcgaactctgagcccaagaagcgaagctgtacc 6180
|||||
Db 6121 TTGACAAGCAGCATGTTCCGAAAGGCGACCTGTAGGCCCAAGACCAAGCTGTACC 6180
QY 6181 tagaagaaccagaccgcgggacctcaagctgctgtctgcgagaggagctgcctgtgtcc 6240
|||||
Db 6181 TAGAGAACCAGTACCCGGGCTCTCACTGTCTGTCTGCGAGGAGTCTGCGCTGTGTC 6240
QY 6241 ccactggtgtgtggtgagcacaactgactgtggtggtggtggtggtggtggtggtcc 6300
|||||
Db 6241 CCACGTGGGTGTGSGGTGACACACTGTACTGTGGGCTGGGCTGTGCTGTGTC 6300
QY 6301 catctgtgtgcatgtgtgtctctctctctctctctctctctctctctctctctctcc 6360
|||||
Db 6301 CACTGTGTGCACTGATGCT 6360
QY 6361 aaaaaaaaaa 6372
|||||
Db 6361 AAAAAAAAAA 6372

RESULT 2
LOCUS AB037816 4886 bp mRNA linear PRI 14-MAR-2000
DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.
ACCESSION AB037816
VERSION AB037816.1 GI:7243170
KEYWORDS
SOURCE Homo sapiens brain cDNA to mRNA, clone.lib:PbluescriptII SK plus
clone:hj07927.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Nagase,T., Kikuno,R., Ishikawa,K.I., Hiroseawa,M. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XVI. The complete sequences of 150 new cDNA clones from brain which
code for large proteins in vitro
JOURNAL DNA Res. 7 (1), 65-73 (2000)
MEDLINE 20181126
REFERENCE 2 (bases 1 to 4886)
AUTHORS Ohara,O., Nagase,T. and Kikuno,R.
TITLE Direct Submission
JOURNAL Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:odnainfo@kazusa.or.jp,
URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
source
1. 4886
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hj07927"
/tissue_type="brain"
/clone_lib="PbluescriptII SK plus"
1. .4884
gene

CDS
/gene="KIAA1395"
<1..>4884
/gene="KIAA1395"
/note="Start codon is not identified."
/codon_start=1
/product="KIAA1395 protein"
/protein_id="BA92633.1"
/db_xref="GI:7243171"
/translation="DARTMAASERRAFKINRTVAAYEYKOVSRERSGPHSRRCSSGLVPLEVEVEDEVEDLSPDPAEPLADLVEFPADDELELLOPREDTPEGIPKDELDVORAVAYVEDMYIVRRROYLSAASPYTTOREKGLPROVEDDASGDRSGSGPDSNDSSRGSPEDTPRSSGASITPDLNLAADSLPDLRLAPEDVDRNETLRQHRPRLTLTPRPDDEDAVERCSREPRHRGPRLTKLSITFEIETIPGILALDVERKKITSENFTELDNSDKGLRKHGTLRLSTLRAITSVTYPSDIEFLVTKLEKVLQGDIDSECEPVYALVVDPAKKELEKRLAEQFRLGRYRMPFAMTVHLANTIVSSAGQLRDSDESGERPMATDRRRGPDSSGDDCSFSGRPATLVTFNFKOERLSDLEDFKLAMRRSSILRLRLVTAOLKIDISPAEENPHCSPELLIKRYPDRGRPTKEILEEPREYVAPHTSYRLNLYVPSLNEFSROGSVYNLAARVOYMGDEPSOALPYIRKSSCSSEPTREAFTPVYHKKSPFEYEPKLIHPACTEENHLFTFTYHSCOPRGTLALTPVGTFTPILOHRLRPRGPRCLASVDOPPSISVLTTPVALPGMKWVDGKGVSELTVAASVHPDPTLDKFTLVHLEEGAPFRUKDVLSEGNVEQLRASLALRLASPEPLVASHVLAFLKLVLRPILISGOIVNLGRGAFEMAHVSVLVRSLLEAADARHCQPLAAVYHAFRLGTEPSLDGAPPVTOAATLARGSPASLYLARSKSISSNPDLAAPGSVDEVSRLASKILHEELALOMVSSAVREAILLOHAFEPOLMKSMALHLLGORLDPKRLPEPRPLDILALVGSGLGVITTRHKDVELAHLNLSAFLPSLDLSLYDGFVPSIYRAIKOYATRLDSSPAPALTLRLMEFTRLCSHSHVYTLNLPCCPLSPSPSPSPSVSTISOSTFSOAPDPKVTSMFELSGPPOQFLGLLTELALALEPAGAFELHKAISVSDTLCHGDHPRYAEFAVRAVAELVPLSLIARDLPLHDPABGPGORSLASVSDTLEGBDLAGTINPVSAMAIAGGPLAPGSRASISOGPTARACALSAESSRLLACVLWVKNTEPALORWADTLPLPOLGRILDLIYLICIAFEYKAKKAFERISILFEKSLDKARLEAIIIGTIGARQEMVRRRESRPNPNPNVRRKSVHMKOTSPRVKTDNEMEHLEVEGNLATEASLVLDLETLIVQVWLSFARSVYAVLKYLYLSGNSQSLFLOHGLAORALVASKPELLEDETLCADCLRLRHGSRISTITHASASLYLMRONEFIGNHFAVMQVYMSLSLVGTQONSEHRLRSKLTITVAEEDMGGRDSTFAEOYODLFENLHLITDPTVKMKHEQDEPEMLIDIMYLRARGOQSPDLRLTLWNMAGKHAEIGNAEEAOCMVHAA"

BASE COUNT 941 a 1605 c 1425 g 915 t
ORIGIN

Query Match 76.6%; Score 4882.8; DB 9; Length 4886;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4884; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 gacgcgagaccatgctgctccgagcgccgctcgcgcaaacatcaagagc 60
|||||
Db 1 GACGCGAGACCATGCTGCTCCGAGCGCGCGCTTCGCGCAAGATCAACAGACG 60
QY 61 gtgagcgagaggtgaggaagcagtgctccgggaacgagtgctcccccaccatcagc 120
|||||
Db 61 GTGGCGGAGAGCTCGGAGAGAGGTGTCGGGGAACGACATGCTCCCCCATCTCAGC 120
QY 121 aagcgctgcagcagctccctgggggtcccaactgactgaagtgtgcagccctgagcttt 180
|||||
Db 121 AGGCGCTGAGAGAGCTCCCTGGGGGTCCCACTGACTGAAGTTGTCAGACCCCTGACTTT 180
QY 181 gagagtgactctgagcccgagccacagatgctgagcccgagccctcaagagctgta 240
|||||
Db 181 GAGAGTGACTTCTGAGCGCGCACAGATCTGAGCCCGGCGCCCTTCAGGAGCTTGTA 240
QY 241 gattcccaagctgactgagctgtgctgctgagcccgaggagatgcggaacagag 300
|||||
Db 241 GAATTCACAGCTGATGACTTGAGCTGCTGCTGCAAGCCCGGGAATGCCGACACGAG 300
QY 301 cccggagatcccaagaagatgaaatgtagcccaagtgtaggagcccgagtgagagatgat 360
|||||
Db 301 CCGGGAGATCCCAAGAGATGAATACTGATGCCCGAGGAGGCGCGCTGGAATGAT 360
QY 361 attgagagctgggtcatgtgtccacagaagatatacgtatcccgagtgtgaggaataagcccc 420
|||||
Db 361 ATTGAGAGCTGGGTCAATGTCACAGAAAGTATCAGTACGTGAGTGCAGCATACAGCCCC 420

Qy	421	gtcaaccaagacacacacgcggagagcgaacaaagggcctcccccgcaggtcttttgagcag	480
Db	421	gtccaccacacacacacagcggagagcagacgaagaggcctccccccgacggttttgacacag	480
Qy	481	gagtcctcttgaaacagagaggtccggcccttaagagactgaatagactcccgctgctcg	540
Db	481	gattgcttcttgagacacagagaggtccggcccttgaggacttgaaatgactctccggcggtccg	540
Qy	541	ggctcccccggaaagacaccccctcgaagcagtggtgctcttaacatctctgcaccttgagac	600
Db	541	ggctccccccggaagacaccccctcgaaagcagtggtgctcttaacatctctgcaccttgagac	600
Qy	601	ctggagagctgactcaattgtgtgcctctctgtctagaagcggcgccacagaagatgtgac	660
Db	601	cttggcagctgactcaattgtgtgccctctctgtctagaagcggcgccacagaagatgtgac	660
Qy	661	cggcgaatgaaacccctcgaagcagcagcggcccgcccgctgcacccctcctacccg	720
Db	661	cggcggcaatgaaaccccttggacggcagcagcggcccgcccgctgcacccctcctgacccg	720
Qy	721	gcaectgacgagatgaaagccgttggaacgctgttagccgcacaaagccaccccgagacac	780
Db	721	gcacctgaaagagatgaaagcccgctggaaagcgtgtagccgcacaaagccaccccgagacac	780
Qy	781	tttgagcaagaagatcttgtaagagttctgcgtctcaagtctgaatgtgaattgaagccc	840
Db	781	ttttggacaaagatcttggctcaagttctgtcgtctcaagttgcaagattgaaattgagccc	840
Qy	841	atccttggaatcttgctcctctatgatgtcgcgagaaagaaagaagatctcgagaaacttc	900
Db	841	attcttgagatcttggtcgtctgatgatgtggggagaaagaaagatcttcggagaaacttc	900
Qy	901	tacttgactctgaactcgtgactgcataatgaaggagctgtctcggcctcatgagcaaccaact	960
Db	901	tacttgactctgaactcgtgactgcataatgaaggagctgtctcggcctcatgagcaaccaact	960
Qy	961	gcatctccacccctgcgcgctctgcacatctctctcttgacccaacccctcacctgaactc	1020
Db	961	ggcattctccacccctgcgcgctctgcacatctctctcttgacccaacccctcacctgaactc	1020
Qy	1021	ttccctggtcatcaagtgtgagaaggtgcttcacgaagggagacacagtgagtgctgtgag	1080
Db	1021	tttccctggtcatcaagtttgagaaaggtgcttcacgaagggagacatcagtgagtgctgtgag	1080
Qy	1081	ccttactagtggttgaaagaagtgtagacaagccaagaacaaagaagactlagagaagctg	1140
Db	1081	cctttacatggtggttgaaagaagtgtagacacgacccaagaacaaagagacttagagaagctg	1140
Qy	1141	cgccctgagcgcgagcagatctcgaaccggccttgccgcgtctacccgaatggcctctgcgcg	1200
Db	1141	cgccctgagcgcgcgagcagatctcgaaccggccttgccgcgtctacccgaatggcctctgcgcg	1200
Qy	1201	aegcgctgtgcaacttggccaacatcgttgagcagcgtctggaactgtgaaccggagactctac	1260
Db	1201	acggcgctgtgcaacttggccaacatcgttgagcagcgtctggaactgtgaaccggagactctac	1260
Qy	1261	tcgagagcgcgagcgcccgccaagctctggaacagacccgcgcgtctcgaggccccagagacgg	1320
Db	1261	tccgagagcgcgagcgcccgccaagctctggaacagacccgcgcgtctcgaggccccagagacgg	1320
Qy	1321	gcgagttagtgagagcagcgcctgcaactctctcgtgctcttcggtccagcgcaagtgtaactg	1380
Db	1321	gcgagttagtgagagcagcgcctgcaactctctcgtgctcttcggtccagcgcaagtgtaactg	1380
Qy	1381	acaaactctttaaagcagaagagcttgagcgaactgaatgacaaagaaactctcaagttcctg	1440
Db	1381	acaaactctttaaagcagagagagcttgagcgaactgaatgacaaagaaactctcaagttcctg	1440
Qy	1441	gctaacactgagcgcccgctgctccctgcgtcgcggtgaactagctctgtgactgtccacgtc	1500
Db	1441	gctaacactgagcgcccgctgctccctgcgtcgcggtgaactagctctgtgactgtccacgtc	1500
Qy	1501	aagatcacactttctcggctccttgaaatccccaactctgcctctccccctgagctgctt	1560

Db	1501	AAATGACATTTCGCGCTCTCGAATAATCCCACTTTCGCTTCGCCGTGAGCTCTT	1560
QY	1561	catataagcccttaeccggaaacccaagggccgcccacaaaggagatctctggagtlcccc	1620
Db	1561	CATATCAACCCCTTACCGGACCCCGAGGGCCGGCCCAACGAAGAGATCTCGAGTTCCCC	1620
QY	1621	gcccggaagttatagcccccatagaccagctcaaggagaaacttgtagtgaacccgaac	1680
Db	1621	GCCCGGAAGTTATAGCCCCCATATCAACGAGTACAGGAACCTGCTGATCTGTATCCCGAC	1680
QY	1681	agccttaacttaagcagccgcagggatctcgtagcgaactctgtgtgagtcgagtaac	1740
Db	1681	AGCCTCAACTTACGACGCCCGACGGGCTCGTGCGAACCTTCTGTGGATGCAATAC	1740
QY	1741	atgacagggagagaccccaagcagcctctgcggtatctttggcaagtcagctgaagt	1800
Db	1741	ATGACAGGGAGAGACCCGACGCAAGCTCTCCGCTCATCTTTGGCAAGTCACGCTGACGT	1800
QY	1801	gaatttaacccggagagcccttaacacccggttgctatcacataaagaatcccccagttctac	1860
Db	1801	GAATTTAACCCGGAGGCTTACACGCGGTGTTACTCATATACAAAGTCCCGAGTTTAC	1860
QY	1861	ggaggagttcaagctgcatcttccagcctcgtgtgaaagaaacataccatccgtgttcaac	1920
Db	1861	GAGGAGTTCAAGCTGTCACTTTCACGCTCGGTGACAGAAACATCACCTGCTGTTCAC	1920
QY	1921	tctctacaatgtcaagtgtgccaacccccgcgggagaaacgtcccttgagaaacccgttgagctt	1980
Db	1921	TCTCAACATGTACAGTGTGCACACCCCGGCGCGGACACTGCTTGGAAACACCGTGGCTTT	1980
QY	1981	acttggatcccatctgctgtgcagcacgggcgtgtgaaagccgcccctctgtctccagtg	2040
Db	1981	ACTTGGATTCCTCACTGTGTGCACAAAGCGCGCTGAGACCGGCGCTTGTCTCCAGTG	2040
QY	2041	tctgtggaaacagccgcgcgcacgactatccgtgtcaaacccgagatgggttcgagccggc	2100
Db	2041	TCTGTGAACACGCGCGCGCCGACGATATTCGCTGTACACCCGAGTGTGGCTTCCGAG	2100
QY	2101	atggcgtgggttggaaaggtlcaaaagggcggtttgaagtgtgagatcacaagccggttccct	2160
Db	2101	ATGGCCTGGGTGGAGCGGTTCACAAAGGCGGTTCAGTGTGAGACTCACGGCGGTGCTCT	2160
QY	2161	gtgacacccccagggacccctacccctgggaaattcttaaccctgtgtgcaagtcctggagag	2220
Db	2161	GTCACACCCCAAGACCCCTTACTGTGCACAAATTTCTTACCTGTGTGCAGTCTTGAAGAG	2220
QY	2221	ggagccttcccatctccgggtccaagagacactgtgtcgtgagggaggaacgtgtgagcaagag	2280
Db	2221	GGAAGCTTCCCATTCGCGGCTCAAGGACACTGTGCTGAGAGGGCAACGTGAGACAGAG	2280
QY	2281	ctggggggcagttcttgcaacatctgcgctctgcccagccccgaacccctgtgtgacctctcc	2340
Db	2281	CTGGGGGCAAGTCTTGTGAAGACAGCCGCTGGCCACCCCGAACCCTTGTGGCTTCTCC	2340
QY	2341	caccacgtctctgaaagaagcctgtggtgtcgtgtgcatcagaagccccgaatcaatgtgagcag	2400
Db	2341	CACCACGTGCTGACAAGCTCTGTGCTTCTGTGATCAAGGCCCCCGATCATAGTGGCCAG	2400
QY	2401	atgttgaaccttgggcggttggagcctttgaagaaatgggccaatagtcacagcctgtgttcaac	2460
Db	2401	ATTGTGAACCTGGGCGGTGGAGGCTTTGAAGCAATGGCCATGTAGTCACCTTGTTTCAAC	2460
QY	2461	cgagagccttggagcgaacccaagagatgcgcgcygtcactgtgccaaagctgtgctctcaagtc	2520
Db	2461	CGGAGCCTGGAGGCAACCCAGAGATGCCCGCGGTCACTGCCACAGCTGTGCTCATGCTC	2520
QY	2521	cactaacgcttctgcctctcgtgaactgagcccaagcccccgagatggggccctccagag	2580
Db	2521	CACTAGCCCTTTCGCTTCTCTGGACATAGCCCAAGCTTCCCGGATGGGGCCCTTCCAGTG	2580
QY	2581	acagtgcaagctgcaacatggccccgtgtctctgtgtgcgcccggaagcctctacccctgag	2640

Db 2581 ACAAGTCAGGCTGCGCACTAGGCGCGTGGCTGTGCGCGCCCGCAAGCCTTACTGGCG 2640
QY 2641 cgttccaagaatcaagcaagcaaaccttgaactctgcctgagcccttgatctgtgagat 2700
Db 2641 CGTTCCAAAGAGCTAGAGCAGCAACCTGTGACTGTGGCGCTGGCTGTGGAT 2700
QY 2701 gacgaagttcccgatcctctgacgaagcgtgtcttcaagaagaagcttgatctgacgtg 2760
Db 2701 GACGAGGTTTCCCGCATCTGGCGCACAGCTGCTTCACAGGAGACTGGCTGTGAGTGG 2760
QY 2761 gtgtatcagacgagtgcgttaacgagagcaactcccaagcagcgtgtctcttccag 2820
Db 2761 GTGTATCAGACGAGTGCCTTACCGAGGCGCATCTTCACAGCAGCGCTGTCTTCCAG 2820
QY 2821 ctcatagtgtaagaagtaaggcgtgacgtctgtctgtgacgaagcgtatagacacacccgc 2880
Db 2821 CTCATGTTGAAGTATGTAGCGGTGACCTGTGCTTGGCGACGACTAGACACACCCCGC 2880
QY 2881 aagctgcgtcttcacgaagcgtctctctgacgaacataactgctgtgtggtctgtggc 2940
Db 2881 AAGCTGCGTTCCTCCCGAGCGTTCGAGGACATCATGCTTGGTGGCTGTGGGCG 2940
QY 2941 ctggaagtcatacccggtgtccaacaagatgtggaagctgagcgaacactcaagccagc 3000
Db 2941 CTGGAGGTATCATCCCGTGTCCACAAGGATGTGTGAGCTGGCGGAGACCTCAACGCCAGC 3000
QY 3001 ctggtcttctctcgaagtgaaactctgtccctgtgtgagccggggtctgtcttcagcctg 3060
Db 3001 CTGGCTTCTTCTCATGATGACCTTGTCTCCCTGTGGAGCGGGGCTTGTCTTCAAGCCTG 3060
QY 3061 gtccggggccacatacaagaaggttgagcagcggtctccagtgctcccttaataccaagc 3120
Db 3061 GTCCGGGCGCCACTACAAAGAGGTGGCCACGGCGTCCACTGCTCCCTTAATCCACAGCGC 3120
QY 3121 ctgtctgaacctgagcattggaatcacccgcatcctgtgtgagccaagcaactacgtgacc 3180
Db 3121 CTGTCTGACCCCTGTGCGATGGAATTACACCGCATTCGTGTGAGCGCAGGACGACTACGTGAGAC 3180
QY 3181 ctcaacctccctgtctgccccctgttaacctcaagctgagctgccccctctgtgtctcc 3240
Db 3181 CTCACCTCCCTGCTGCGCCCTGTACCTTCACAGCTCCCTCCCTCTGTGTGTCTCC 3240
QY 3241 accaactcccaaggtccacactctctcagccaagccccgagcccccaaggttgacagcatg 3300
Db 3241 ACCAAGCTCCAGAGCTCCACTTCTCCAGCCAAAGCCCCGAGACCCCAAGGTGACCAAGCATG 3300
QY 3301 ttcgaactgagtgagcaacttccggcagcagcaacttctctagctggggtcctctgtcagcag 3360
Db 3301 TTGGAAGTGAAGTGAAGCAATTCGGGAGAGCACTTCTAGCTGGGCTCTGCTGAGAGGAG 3360
QY 3361 ctggagctggccctcgaactgaaggttggaagggcatctcctgttgacaaagaagccatc 3420
Db 3361 CTGGAGCTGGCCCTTGCAACCTGAGGCTGGAAGGGGCAATTCCTGTGTGCAAGAAAGGCATC 3420
QY 3421 agtgcgtgtcaagcgtctgatagttgacatagaactgacccccgcgtacagccgaagccat 3480
Db 3421 AGTGCCTGTGACAGCTGTGATGTGGCATGTGACACTGACCCCGCTACGCCGAGGCACT 3480
QY 3481 gtgaaggtcgtgtgacgaagctgtactcgtgcaactgcttgcattgacagggataacttg 3540
Db 3481 GTGAAGGCTGTGTGGCCGACAGCTGTACTGCACTCTTTCATTGACAGGAGTACTGCTTG 3540
QY 3541 ccaagcgtgcatgacttctgaagggccaggttcagcgggtcaagactggtccctcaatgctt 3600
Db 3541 CCAGCGCTGTATGACTTGTGTGAGGGCCCAAGGTACAGCGGTCAAGAGTGGCTCTCAATGCTT 3600
QY 3601 gactcagacacayaagcgaaagggagacattgcggttaccatcaaacccctctgtgacatg 3660
Db 3601 GACTCAGACACAAAGGCGAAGGGGACATTTGGGGTACCATCAACCCCTCTGTGGCGATG 3660
QY 3661 gcatctgctgtgtgccccctagccccctgtgctccgggacagcatctccaggggcacaca 3720
Db 3661 GCATTTGCTGTGTGGCCCTTAGCCCTTGCGCTCCGGGCGACGATCTCCAGGGGCGCACCA 3720

QY 3721 aaggtctctgcgcaagcgtgtgtccctctctgtgtgaagtcgaagccggaactgtgtggtgt 3780
Db 3721 AAGGCTTCTCCGCGAGCGCTGTGCCCTTCTGTGACTGTAGCAAGCCGAGACTGCTGTGGCGTGT 3780
QY 3781 gtgctgtgtgtgtctgaaaaaacacccgaagccggcgctccttgaaagcgttgagccacttgac 3840
Db 3781 GTGCTGTGTGTGTGTAAAAAACCCGAGCCGGCGCTCTGTGAGCGCCTGTGGCCACTGTGACTG 3840
QY 3841 aacttccccagctggagcgtctgttgactgtgtgaacttgccttgctgctgtgcttgag 3900
Db 3841 AACCTCCCGCCAGCTGGAGCGTGTGTGACTGTGTACTTGTGCTGTGGCTGTGGCTTGTGAG 3900
QY 3901 tacaaggggaaaaagccttgaacgcatcaaacgctccatcaaaatctctgagat 3960
Db 3901 TACAAGGGGAAAAAGCCTTTGAACGCATCAACAGCCTCAATCAAAAATCTCTGGAT 3960
QY 3961 atgaagcgcgcgctaaaggggaagccattctgtgtgtccatctggaagcttgaaagaatgtgt 4020
Db 3961 ATGAAGCGCGCGCTTAGAGGAAGCCATTCTGGGTACATCGAGCTCGACAAAGAAATGTGT 4020
QY 4021 cggcgaaagctgagagagagccgcttgggaatccagaagaaagtgtgcgttgcggaagagc 4080
Db 4021 CGGCGAAGTGTGAGAGAGCGCCGTTTGGGATTCAGAGAAACGTCGCGCGGAAGAGC 4080
QY 4081 gtcaacaacttgaaagcaaacctcagaccgctgtggaacaagaaccaagatgaatggaacac 4140
Db 4081 GTCAACACTGTGAAGCAAACTCAGACCGCGTGACAAAGACCAAGGATGAAGATGGAACAC 4140
QY 4141 gaagcctgtgtgaaaggaagccttgcaacggaggcaagccttaagtgtgtcttggaacacgt 4200
Db 4141 GAGGCTTGTGTGAAGGGAAGCACTTGCAACCCAGGACCAAGCCTTAAGTGTCTGTGGACACCTG 4200
QY 4201 gagaatcactgtcagacggtgatagtcttcagaagccccgggagagagcgtcttgtygggagatg 4260
Db 4201 GAGATCATCTGTGACAGCGGTGATGCTTTCAAGACCCGGGAGACGTCCTTGGGGGCGAGTG 4260
QY 4261 ctgaaggtgtgtgtgtacagcttgaggcagtggtgccagagtgctctctcttgcaagcatgagc 4320
Db 4261 CTGAAGTGTGTGTGTATAGCTGTGGGCAAGTCCAGAGTGCCTCTTCTTGTGAGCAATGGC 4320
QY 4321 ctggacacccagagagggccctgtgttccaagtctcccgagcgtgctgttctgagaggaacag 4380
Db 4321 CTGGCACACCAAGAGGCGCCTGTGTCCAAAGTCCCGGAGCTGCTGTTCAGAGAGGACAGC 4380
QY 4381 gaactgtgtgcgaactgtgtcctgaaggtcctctacgacactgtgtgagcgcgcatcagacc 4440
Db 4381 GAGCTGTGTGCGGACCTGTGTGCTGAGGCTCTTACGACACTGTGGCACGCCCATCAGCACCC 4440
QY 4441 atccgcaagcagcgaagcgctctgactgaactgtctatggaacgaagaacttcagagatcgac 4500
Db 4441 ATCCGACGCGCAGCGCGAGCGCTGTGCTGTACTGTCTATGCGACAGAACTTGAGATGGC 4500
QY 4501 cacaacttgcgcgtgtgaaagatgcaggltcaacatgtctctctcctgcctgggttgaggagc 4560
Db 4501 CACAAGCTTGGCCCGGTGAAGATGCAGGTACATGTCTGTCTGTGCTGTGGGGAGCG 4560
QY 4561 acgcaagaacttcagtgaaagagcacttgagcgttcaactcaaaacactcactaactgtct 4620
Db 4561 ACGCAGAACTTCAAGTGAAGAGCACCTGTGAGCTGTCACTCAAAAACATCTCACCTATGCT 4620
QY 4621 gagagagacatgagggctgtgcgagacagcaacttgcaagagcaggtgtccagagccgtgattc 4680
Db 4621 GAGGAGGACATGGGGCTGTGGGACAGCACTTGTGCAAGGAGGCTGTGAGGACCTGATTTTC 4680
QY 4681 aaactgtcacatgataccttgacgagacaggttgaaagatgaagaaacccaagaggaacccgtgag 4740
Db 4681 AACGTGCACATGATTCGTGACGAGACAGGTGAAGATGAAGGAACCAAGGAGGACCTGTGAG 4740
QY 4741 atgtctcatcgacctcaatgataagaattgcccggggtctaccagggctcaacggagacttcgg 4800
Db 4741 ATGTCTCATCGACCTCATATGTATCAGAAATGGCCGGGGCTTACCAAGGGCTTACCGGACCTTGG 4800

OY	4801	ctgacctggctgcagaacatggccgggaagcagcgaggcttggcaaccacgccgagggcc	4860
DB	4801	ctgacctggttcacaaacttgcccg6gaagcacgcgagcttg6gcacacccgagggcc	4860
OY	4861	gccacgatgcatggtgcacgcg9gcgcg	4886
Db	4861	gccacgatgcatggtgcacgcg9gcgcg	4886
RESULT	3		
LOCUS	BC008335	2825 bp	mRNA linear PRI 12-JUL-2001
DEFINITION	Homo sapiens, clone IMAGE:3506202,	mRNA,	partial cds.
ACCESSION	BC008335		
VERSION	BC008335.1	GI:14249910	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 2825)		
AUTHORS	Strausberg, R.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.ncl.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbsc@mail.nih.gov Tissue Procurement: DCMD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLP) DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada info@cgsc.bc.ca Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ken Guin, Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Scheidt, Duane Smalins, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Series: IRAL Private: 23 Row: d Column: 12. Location/Qualifiers		
SOURCE	1..2825		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3506202"		
	/tissue_type="Kidney, renal cell adenocarcinoma"		
	/clone_1lb="NH_MGC_14"		
	/lab_host="DH10B-R"		
	/note="Vector: pOTB7"		
	<1..2585		
CDS	/codon_start=3		
	/product="unknown (protein for IMAGE:3506202)"		
	/protein_id="AAH08335.1"		
	/db_xref="GI:14249911"		
	/translation="QSRSLASMLDSDTEGEGDIAGTIINPSVAMAIAGPLAPGSRAISOPPTSRAGKCAFLAESSTTLIACVIAMTKNEPALLOWAWDLTPOLGRILLIDLYLCIAAPEYGGKAFFERINSITLFEKKSIDMKARLEELIGTGGAODEVRYSRRESPPGNPEVBRPKSYTHHKOTSDDRYDKTRDEKEHEVALVEGNATLASVLVDLTLETIVQWLSARESVGLAVLKVVLLSUGSAOSAFLOLGATQRALVSKRPPELLFEEDELCAADCLRLRLRCGSRISTIRTHASASLYLMKRONETIGHNFARKVMQVTMSISLVGTGONFS EHLRLRSKITLIYYAEEDMGIRDSTPAEOGYODIPNHHMLITGVYKRKEHQEPENMI I DLMYLRARGYGSPDRLRTJLWOMACKHAELGHNAEAOOMVAALAVLEYIALLDIH RHLPVCVSFONSNNVLESATSDDISIDSGDEPGSGKRFTELGLVGLLEQAGAGFT MGSLVANVENVYKNILPILEANHDYRKLAAYHRKLODAFTIKHSSGMWRVGTYFR VFTYGAHFQDLDEQEFVYKEPSTIKLAIESHRIEEFYTFEEKGDVVELIKDSPNPVKS		

[illegible]

OY	4472	tgtctatgcgcagaaacttcgagatcggccacaaactttggccggtgltgaagtgcagttca	4531
Db	901	tgcctatgcgcagaaacttcgagatcggccacaaactttggccggtgltgaagtgcagttca	960
OY	4532	ccaatctctctcgcgccttcgtgtgggagagcgcagaaacttcagtgaagaacccctgcac	4591
Db	961	ccaatgctctctcgtgtgtcccttcgtgtgggagagcgcagaaacttcagtgaagaacccctgcac	1020
OY	4592	gttcaactcaaaaacccatctctcaacctatgtctaggaggaacatggggctgcggagcaacct	4651
Db	1021	gtttcaactcaaaaacccatctctcaacctatgtctaggaggaacatggggctgcggagcaacct	1080
OY	4652	tgcgcagagagagttccaggaacccgatggttcaacctcaactgtacttcgtacgcagcgtgtga	4711
Db	1081	tgcgcagagagagttccaggaacccgatggttcaacctgtacttcgtacgcagcagcgtgtga	1140
OY	4712	agatgaaggaacacacagagggagacctgagatgtcatctgacctcalgtiacagaatlyccc	4771
Db	1141	agatgaaggaacacacagagggagacctgagatgtcatctgacctcalgtiacagaatlyccc	1200
OY	4772	ggggcttaacaaagtctaacggagaccttcggctaacctgggttggaagaacatggcccggaac	4831
Db	1201	ggggcttaacaaaggctaacggagaccttcggctaacctgggttggaagaacatggcccggaac	1260
OY	4832	acgcgagagcttggtgcacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	4891
Db	1261	acgcggagagcttggtgcacacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1320
OY	4892	tggctgagtaacctgcgccttcgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	4951
Db	1321	tggcttaagagtaacctgcgccttcgtctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc	1380
OY	4952	tccagagacatcatccaacgctgctagagagagtcgcgcacatctccgcagacatccgtcgc	5011
Db	1381	tccagagacatcatccaacgctgctagagagagtcgcgcacatctccgcagacatccgtcgcgc	1440
OY	5012	ccgagcagagagaggtcttcgtctccgcggaaagcaacttaactgagctcggggctggtatggcttc	5071
Db	1441	ccgacagagagaggtcttcgtctccgcggaaagcaacttaactgagctcggggctggtatggcttc	1500
OY	5072	tggaaacagagagccgcgttacttacaatggcgcgctctcagagcggttgaaatgaagttct	5131
Db	1501	tggaaacagagagccgcgttacttacaatggcgcgcttactcagagcggttgaaatgaagttct	1560
OY	5132	acaagaacctcatcccatccctcgtggaagcccaacccgttaactacaagaagctctggccggtgtc	5191
Db	1561	acaagaacctcatcccatccctcgtggaagcccaacccgttaactacaagaagctctggccggtgtc	1620
OY	5192	acggcaaacatgcagagagagcccttacaagaatcatgcacaaagatcccgagcttggaagcgcg	5251
Db	1621	acggcaaacatgcagagagagcccttacaagaatcatgcacaaagatcccgagcttggaagcgcg	1680
OY	5252	tgttcggagagctatlttcgcgctggggttcttaccggcgcgaacttcggttgaacttgatgagc	5311
Db	1681	tgttcggagagctatlttcgcgctggggttcttaccggcgcgaacttcggttgaacttgatgagc	1740
OY	5312	aggagatttggtgataaagagagcaatcagatcagagagctgggaagagatctcaaccgggtctg	5371
Db	1741	aggagatttggtgataaagagagcaatcagatcagagagctgggaagagatctcaaccgggtctg	1800
OY	5372	aggagttctcaacgcgagagatattggcgacgacgtctgttgatattacaagaactcttacc	5431
Db	1801	aggagttctcaacgcgagagatattggcgacgacgtctgttgatattacaagaactcttacc	1860
OY	5432	ctgttgagaagaatccaagcttgaactacaagaagcctatcatccagatcaagtatgtggaac	5491
Db	1861	ctgttgagaagaatccaagcttgaactacaagaagcctatcatccagatcaagtatgtggaac	1920
OY	5492	cgtaactttgatactacagatcgaagtcgaagacccggtgtgaacttattggaacgcgaactatggc	5551
Db	1921	cgtaacttttgatgaactacagatcgaagtcgaagacccggtgtgaacttattggaacgcgaactatggc	1980
OY	5552	tttcgcacatttccttctgcacgcggttcaacgcgcgaatlyggcgcgcacacgcggagctgc	5611

[illegible]

source 1. .2149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="preliminary human CLASP-7 cDNA sequence"
CDS <2..1933
/note="human CLASP-7"
/codon_start=1
/protein_id="CAC43732.1"
/db_xref="GI:14598039"
/translation="LSFAESVYLGLLVLYLSGSASALFLHGLATVORALYSKPE
ELLFEDPEEDCADLCLRLRHGCSIRITRTHASLYLIMKONFEIHNARVKMOY
TMSLSLVTTONFSEELHRSNLTITLYEEDMCLRSTAEQVQDLMFNHMLITD
TVKKEHODEPEMLIDLMWRIARGVQSPDLRLTWLQNMAGHAEIENHDAACOMVH
AALVAEYLALLEDHRHLPLVGCVSFQNTISSVLESAISDILSPDEGFCGKHTE
LGLVLELQAAEYFTMGGLYEAVNEVYNKLLPILEAHRDYKRLAIVHCKLOEAFKTM
HOSGMEVFEGYFRVGFYGAHFDLDOEFVYKPSITTKIALEISHRELYTERFGD
DYVELIKDSYPIYDKSLDSOKAYIOITTYEPEFTYELKDRVTFDRNGLRTLEFCT
PPTPPGRAGNELPEQHKRKTLLSTIDHAFYIKTRIRVGHREYVLTPEVAIEDMOK
TRELAFTEQDDPPDAKMLQWLVQSVPTVNOGPLEVAQVLAETPEDPKLFRHNNK
RLCFDFCKCEDALRNKALIGSPDKFYHRELEERNYCRLEALQPLTLQRLPQLMAP
TPPGLRNSINRASFRKADL"

BASE COUNT 468 a 650 c 625 g 406 t
ORIGIN

Query Match 33.6%; Score 2140; DB 6; Length 2149;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2143; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

4225 cttcagaagcccgagagagcgtcttgggggcaagtcgtgaaggttctgtctgaagcctg 4284
2 CTTTCAGAAAGCCCGGAGAGCCTTTGGGGATTGCTGAAGTGTCTGTACAGCCTG 61
4285 ggcagtcgccagagatgcctctctcttcagcaatgcctggccaccagagagccctgtg 4344
62 GCGAGTCCCAAGAGTCCCTTCTTTCAGCATGGCCTGGCCACCACAGAGGCCCTTTGT 121
4345 tccaagttcccgagcgtgtgttctgaagagagcaagagcgtgtgtgcgaacctgtgctg 4404
122 TCCAAAGTTCCCGAGAGTGTGTTGAGAGAGGACACGAGAGCTGTGCCGACCTGTGCTG 181
4405 aggcctctcagcaactgtgagcagcgaatcagcaacatccgacagcagcagccctcg 4464
182 AGGCTCTACGACACTGAGGAGCGGACATACGACATCCGACGACGACGACGCTCG 241
4465 ctgtaccctgcatatgcagcagaactctgaagatcgccaaacttgcocgtgtgaagatg 4524
242 CTGTACTGTCTCATCGACAGAACTTCAGATCGGCCCAACTTGGCCGCTGTGAAGATG 301
4525 caggtcaccatgtctctctcgttccctgtgtgggagcagcagaacttcaatggaagagcac 4584
302 CAGGTACCATGTCTCTCGTCCCTGGTGGGAGACGACGAACTTCAGTGAAGAGCAC 361
4585 ctgcgaagttcaactcaaaaacatccatcctaatactgaagagagacatgaggtctgcgagac 4644
362 CTGCACAGTTACTCAAAACATCTCCTACCTATGCTGAGGAGGACATGGGCTCGGGAC 421
4645 agcaccttcgagagcaggttcagagacatgattcaaacctgcacatgattccatgaagac 4704
422 AGCACTTCGACAGACAGAGTCCAGAGCTGATGTTCAACCTGCAACATGATCTTACCGAC 481
4705 acggtgaagatgaagaaacaccaaagagacccctgaagatgctcatcgaactcatatgaaga 4764
482 ACGGTGAAGATGAAGAAACACCAAGAGAGACCTTGATGATGCTCATCGACCTCATGTACAGA 541
4765 attcccgagggtaccaggggtaccaggaacttcggtgacactgtgttgcaagaatgagcc 4824
542 ATTGCCGGGGCTACCAAGGCTTCACCGACCTTGCTGACCTGCTGACGACATGAGCC 601
4825 ggaagacacgagagctgagcaaacacgacgagagccgcccagttgcatgtgagcagagcc 4884
602 CGGAAGCAGCGGAGAGTGGGCAACACACGCGGAGGCCGCCAGTGCATGCTGACGCGGCC 661

4885 gccctcgtgagtgagttacatcgtccctgtctcgaaggaaccagcgcaacctgcccgttggtctgc 4944
662 GCCCTCGTGGCTGAATACCTGCGCTGCTCGAGGACACCCGACACTGCGCGTGGCTGCG 721
4945 gttccctccagaacatcatcacaagctgctagaagaggtccgcacatctccagaacatc 5004
722 GTTTCCTTCCAGAACATCTTCATCCAAATGCTGTAGAGGAGTCCGCAATCTCCGAGCAATC 781
5005 ctgtccgcagcagaagaggtctctgtctccgggaagcaacttcaactgaactgtgagctgtga 5064
782 CTGTGCCCCGACAGAGAGAGGCTTCGTCCGGAAGACACTTCACTGAGCTGCGGCTGTGA 841
5085 ggtttgcgggaacacggcagccggtctacttcaacatggggcggtctcagaagcggtgaat 5124
842 GGGTTGCTGGAAACAGGACGCGGCTACTTCAACATGGCGGGGCTCTCGAGGCGGTGAAT 901
5125 gaagttcacaagaacctcatcccatcctgtgaagcccaacgtgaactcaagaagctgtgccc 5184
902 GAGGTCTACAAAGAACCTTCATCCCATCTCGAAGCCCAACGCTACTACAAAGAGCTGGCC 961
5185 gcggtgcacggcnaactgtcagagagccttcaaccaagatcatalgcaaccagagttccggtcg 5244
962 GCGGTGCACGCGCAAACTGCAGAGAGGCTTCACCAAGATCATGACACAGAGTTCGGCTGG 1021
5245 gaagcgtgtctcgagagctatttccgcgtgaggtctcagagggcccaacttcggtgacctg 5304
1022 GAGCGCGTGTTCGAGACGTATTTCCGCTGGGCTCTTACGGCCCACTTGGGTGACCTG 1081
5305 gatgaagagaagtttgttatacaagagcactcatcagaagcttgcagaagactcaaac 5364
1082 GATGAGCAGAGATTTGTGTCAAGAGACCATCATGATCAGAAAGCTGGGAGATCTCACAC 1141
5365 cggctgaagagaattctatacagaagagagatgttgcgaagagctgttgaagatatacaagac 5424
1142 CGGCTGAAGAGGATTTCTACAGAGAGATTTGGGACGACCTCTTGAAGATTATCAAAAGAC 1201
5425 tctaacctctgttggaacaagcttgacttcaaaagagccatactcaagatcaagatg 5484
1202 TCTTACCTCTGGCAAGTCCAAAGCTTGACTCAAAAGGCTTACATCCAGATCACTGAT 1261
5485 gtggaacgtaactttgatactactcagagctcaagagccgggtgaacttctgaacggcaac 5544
1262 GTGGAAACGTAATTTGATATCTACGAGCTCAAGAGCCGGGTGACTTATCAACCGCAAC 1321
5545 tatgggcttcgacatctccctgttctgaacgctgtcaagccggaatggggcgcaacaggg 5604
1322 TATGGGCTTCGACATTTCTGTTCTGCACGCCCTTACGCCGAGTGGGCGCCACACGGG 1381
5605 gaagctgccagagcaacaagaagcgttaagacgctgtctgaacagcagaacgaccttccctac 5664
1382 GAGCTGCCGAGCAACCAAGCGTGAAGCGTGTACGACCGACACGACCTTCCCTTAC 1441
5665 atcaagactgcacatccgttgttgcacacggggaaggaaggttctgaagccaaatggaagt 5724
1442 ATCAAGACTGCAATCCGTGTGTGTCACCCGGAAGAGACGCTGTGCGCAATGGAAGGTG 1501
5725 gccatcgaaggaatctcagaagaagaacacgagagcgttgaccttgcacacggaggaagacca 5784
1502 GCCATGAGAGACATGCAGAAAGAACACGAGAGCTGGCCTTTGCCACCGAGAGAGACCA 1561
5785 ccagatgcataagatgtataagatgtgtctcaagagctcgtgtagggcccaacggtgaacag 5844
1562 CCAAGATGCTAAGATGTGTACAGATGGTGTTCAGGCTCTGTAGGGCCACCGTGAACACG 1621
5845 ggtcccttggaagtgccccaagtglttataagaagatcccggaagacccaagctcttc 5904
1622 GGTCCCTTGGAAGTGGCCCAAGTGTGTTTATGACAGATCCCGAAGAGACCCCAAGCTTTC 1681
5905 cggcatcaaaaataattgcgctctgtcttcaaggaactctcgaagaatgtgagagatgcg 5964
1682 CGGCATCAACAAATTTGGGCTTGTCTTCAAGGACTTCTCAGAAGAAATGTGAGAGATCG 1741
5965 ctgcgaaataaagaccctgtgtggtgcggaaccaaagaggttaacacgctgaactgtgag 6024

Db	Accession	Source	Organism	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
Db	1742	CGCGGAAAAATTAAGCCCTGCTATTGGGCGGACCGAAGAGATACACCGTAGACTGGAG	human.	AX172874	Sequence 1 from Patent WO0142297.	AX172874	AX172874.1	GI:1459711
Qy	6025	cgcaactcctgcyccgcctgcyggaagctctgcagccctctgtctaccagcgcctgccccag	Homo sapiens	6828 bp	DNA	linear	PAT 03-JUL-2001	
Db	1802	CGCAACTCCTGCTCGCGCTGGGAGGCTCTGGAGAGCCCTGCTTACCGCAGCGCTTCCCCAG	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	AX172874				
Qy	6085	ctgatygcaacccacccccccgcctcaaggaaccccttgaaacagagcaatttcggaag	1 (bases 1 to 6828)					
Db	1862	CTGAGGGAACCCACCCACCCCGGCTCAAGGAACCTTTGAACAAGAGATTTCCGAAG	Lu, P., Garman, J.D. and Candia, A.F.					
Qy	6145	gcaagccctctgagcccaagaagaccaaagctgtatccagagaaacagaccgcgcgc	Class-3 transmembrane protein					
Db	1922	GCAGACCTCTGATGCCACCAAGAGCCAAAGCTGTAAGAGAACGACACCGGGCTC	Patent: WO 0142297-A 1 14-JUN-2001;					
Qy	6205	agctgtctgtgctgcgaagggaagctcctccctgtgtgcccactggtggtggaaccaca	Arbor Vita Corporation (US)					
Db	1982	AGCTGCTGTGTGCTGGGAGGGGAGTCTGCCCTGGTGGCCACGAGGCTGTGGGGTGACCA	Location/Qualifiers					
Qy	6265	ctgtactcggggcctgcygagccctcctccctctgtctcccactctgtgtgcaatgctcct	1..6828					
Db	2042	CTGTACTTGGGGCTTGCGCCCTCTGCCCTGTGTCCCACTGTGTGTGCACTGATGCTTCT	/organism="Homo sapiens"					
Qy	6325	ccctttttaaatttaaatggtttttaaagcaaaaaaaaaaaaaa 6372	/db_xref="taxon:9606"					
Db	2102	CCCTTTTAAATTTAAATGCTTTTATAGCAAAAAA 2149	/note="full-length human CLASP-3 cDNA"					
FEATURES	source							
CDS								

[illegible]

[illegible]

Db	3380	TTTACGACGATCAATCAAGCCTTTTGAATGCTGCTACTTTGGTTCTCAAAAATGACG	4039
Qy	3806	agccgcgcctctgcagcgcctggccaactgcactcaactccccagcttggagctgt	3865
Db	4040	ATGAACAGTTCTACAGAGATGGTTTACGAACTCTCACTCTTGGAGTAAACCGGCTAT	4099
Qy	3866	tgcacttgcctgtaaccttgccttgccttgcctttgatacaagggtgaagaccttgac	3925
Db	4100	TAGATCTGCTTTATCTCTGTGTCTTTCGTTTGAGTATTAAGGAAAAAGTTTGAAC	4159
Qy	3926	gcalcaacagctccaatlcoaaaaalctctgatalgaaggcgcgcctlagaggaacca	3985
Db	4160	GAATGATTAAGCTTACGACTTTAAAGAAATCAAAAGACATGAGACAAAGCTTGAAGAGCTA	4219
Qy	3986	ttctgggtaccatcgcagcctgcagaagaatglttcgggaagtctgt	4033
Db	4220	TTCTTGGGACATATAGGGCCAGGCAAGAAATGATGAGGGAAGCCGAGACAGCTCGAGA	4279
Qy	4034	-----agagagaccgttttgggaatccagagaacgtgccttgcctgcggaagagctca	4084
Db	4280	GAACCCATCTGGAAAGTGCTTTGGAAATCAAGAAATTTGAGGTGAGAAAGATATGA	4339
Qy	4085	cacactggaagcaaacctcagaaccgcgttggacaagaaccagaatgaaatgaaaccgag	4144
Db	4340	CTCACTGGCGTCAAAACACAGAAAGACTTGACAAATCAAGACACAGATTTGAACAGAG	4399
Qy	4145	ccttgggtggaaggaaccttgcacacgaggaagccttagtgttctcgcacaactgcgaga	4204
Db	4400	CACGTGATGATGAAACCTGGCTACGAAAGCAAACTATTCATTTAGATACATTAGAGA	4459
Qy	4205	tcatcgtgcagaacgtgatagtcttccaagaaccgcggagagcgtcttggggcagtgctga	4266
Db	4460	TTGTGTTCAACACGTTTCTGTAAACGAAATCAAAAGAGCATTTGGTGAAGTGCTAA	4515
Qy	4265	aggttgcctgtacagccttgcggaagtcccaagagtccctctcttcgcagcatgcctgg	4324
Db	4520	AAGTGTACTACACAGATGGCTGTAAACAAAGTCAGTTATCTACAAACCTGTTTG	4576
Qy	4325	ccaccagaaggcccttgtgtccaaagtlcccgagctgtctgttcgagggagacaacggagc	4388
Db	4580	CTACACAGAGAGCCTTGTTTCAAAAGTTTCTCAACCTTATTTTGAAGAAAGACAGAC	4633
Qy	4385	tglttgcgaccttgccttgcagctccctacagaaccttgcacgcgacacgaacatcc	4444
Db	4640	AGTGTGCTGATTTATGCTCTCAGGCTTCTCGACACACTGTAGACATCGGTGACATAC	4699
Qy	4445	gaacgcacgcacgcgcctgcctgtacctgtcaltgcgaagaaccttcgagatccgcaca	4500
Db	4700	GGTCAACCCCAAGTGCCCTCCCTTACTACTAATGAGGCAAACTTTGAGATTGGGAATA	4755
Qy	4505	acttgcgccgtgtgaagatgtaggttcaacctgtctctctgtccctgcgttggggagacgc	4566
Db	4760	ACTTTGCCAGGGTTAAATGAGGTACCAATGTCACTATCTCTTGTTGGGACATCTTC	4811
Qy	4565	agaacttcaagaagagacactgcagcttcaactcaaaaacatccctcaactatgtcgaag	4621
Db	4820	AGAATTTTAAGAGAAATTTTAAAGAGCTTCTAAAGACTATATTGACATATGCTGAG	4877
Qy	4625	aggacatggggctgcggnagacacacttcgcagagacaggttccagagactgattcaacc	4688
Db	4880	AAGATCTGGAATTGAGGGAACAACATTTCTCTGTACGGTCCAGGATCTGGTTTCAATC	4933
Qy	4685	tgcataatacttcgagggacacgltgaagttgaaggaacacagggaggaaccttcgaatgc	4744
Db	4940	TCCATATATCTTTCTGTATCTGTGAATTAAGAGAAACACAGAGAGATCTCGAATGT	4999
Qy	4745	tcatagacctatgtacagaaatgcgcggggctacagggctcaaccggaacttcggctga	4800
Db	5000	TGATGTATCTATGTACAGAAATTTGCCAAGGTTTACCAAGCTCTGCCAGA---GCGATTGA	5055
Qy	4805	ccttgttgcagaacatggccggggaagcagcgagagcttgcggcaaccacgcgcgagccgc	4866
Db	5057	CTGTGTGGAGACATGGCAGGCAAGCACTTAGAACCAAGCAATTCATGCTGAAGCTGCAC	5111

Qy	4865	atgagcaatggtgcaacgcggccgcccctgttgctgaatccctgcgcccctgcgagaccacg	4924
Db	5117	aggtgttaactccacacacacacacactttgtgctgaattttgacacatttgagagaccga	5176
Qy	4925	ggcaacctgcgcgttggtgcgtgttccctccagaaactctatccaaatgcgtctagaagagt	4964
Db	5177	aaatattcttcttggggattgtttaacatttcagaaatttttcaattatgttttgaagaagt	5236
Qy	4985	ccgcaactccgacagacacatccctgcgcgcgcagaggaagggctcttgctccgggaagact	5044
Db	5237	ctggcggtctacagatgatgtgtgtatctccaatgaagaaggtatctgccttggaanaattct	5296
Qy	5045	tcaatgagctggggctggttaggtgtctgtgaacaggacggctacttcaacatvgggcg	5104
Db	5297	ttactgagtcagacgaacttctgggatttactggamaacacactgcttcttcttgaagctg	5356
Qy	5105	agctctacgagggcggtgaatgaagttctacaagaactctctcccaatccctggaagccacc	5164
Db	5357	gcattgtatgaacgaatttatgaattttacaaactacttatttccattatcattgaagctatgc	5416
Qy	5165	gtgaactacaagaagctggtgcgcggtgtgacgycnaactgtgaagggcttcccaaatga	5224
Db	5417	gggattgcaaaagaactattccacaaattcatgtgtaacttcaaaabacatttcagcaaatgc	5476
Qy	5225	tgcaccagaatccgcgtctggagacgcgtgtctggagacgtattccgcttggtgtctacg	5284
Db	5477	tttcattcaagatgacgtccgtggagcggatgttttgccactatttttgcgtgtggttttatg	5536
Qy	5285	ggccaccattctgggtgaccttgatgaagagagttgtgtacaagaagccatcgatcaaga	5344
Db	5537	gaaccnaatttgcgggattttggatgaacagaaattgtttttacaaaggaccttcacaaacaa	5596
Qy	5345	agcttgcagagatctcacaacgcgcgtgagagagttctcacaacgagaagatttgcgaacg	5404
Db	5597	aaacttgcagagatattctcacagatttggagggattttacggnaaaagaattttgcaagagatg	5656
Qy	5405	tgtgtgagatatacaagaactcttaacctgtgtgacaagtlccaagcttgactacacaaag	5464
Db	5657	tggttgaagtatataaaagactcttaacctgtgtgacaagtlccaagcttgactacacaaag	5716
Qy	5465	octatactcagatacgtatgtggaacgcgtacttggatatacctaagagctcaagaaccggg	5524
Db	5717	catattttcagatttacctatgttgagaccactattcttgacacatnagaatgaagaagacgaa	5776
Qy	5525	tgaactacttgaacggaactatggtgctgcacatctcgttcttgacgcgcgttccagc	5584
Db	5777	tcacctattttgcacaaaatttacaaattcttgcgtgattcatgactgactgacaccttttactt	5836
Qy	5585	cggatvggcgcgcacacgcgggagctgtcccgagacacacaaagcgtatagaagctgtctcaga	5644
Db	5837	tagattggcccgcccaatgggggaacttcattgaacacaaatttcaaaaaggaagcactttctgacta	5896
Qy	5645	ccgacacagccttcccttaatactgaagctcgaatccgcgtgtgtgcacccgggagagagaagg	5704
Db	5897	cgtctcatgcttttctttattatataaaacagggtcaaatgtcactatnaaagaagatgata	5956
Qy	5705	tgctgcagccagatgtaggtgtggccatcgaaagacatgcagaagaagacacgggaactggcc	5764
Db	5957	tctttaacaccaatttgaagttgctatgttagagacatgcacaaaaaagacacagagattggcat	6016
Qy	5765	cttgcacacgagcaggaaccacacagatgctaaagtctaaagatggtgtcttcaggcgtctg	5824
Db	6017	tttgcacacacatcttcagatgtccggcagaccccaaaatncttcaatgtgtaactccagagattcg	6076
Qy	5825	tagggccacacgtggaacacgggttcccttggaggtgtgcccaggtgttlltagcagagatcc	5884
Db	6077	taggcacacacagatgaatcaggggaccttttggaaattggccacagttttctcttgcgaattac	6136
Qy	5885	cggagagacacagccttccgcgcgtacacaaazaatttgcgtctgtcttcaagaactctt	5944
Db	6137	ctatgatcccaaaagcttcttcaacacatcatatataaactgcacgtctgtgttttaaaagatttta	6196

[illegible]

TITLE Direct Submission
JOURNAL Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 152-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp),
URL: <http://www.kazusa.or.jp/huge>, Tel: 81-438-52-3913,
Fax: 81-438-52-3914

BASE COUNT	1312 a	859 c	920 g	1220 t
ORIGIN				

Query Match	23.0%;	Score 1468.2;	DB 9;	Length 4311;
Best Local Similarity	65.1%;	Pred. No. 1.4e-224;		
Matches 2210; Conservative	0;	Mismatches 1158;	Indels 27;	Gaps 2

QY	2713	cgcttcggccagcaagctgcttcaagagagagcgtcgtctcgcagctggtgtgtcgaagcagc	2772
Db	446	CGCTTACCACTTAAAGCTTTTTCAGAGAGGAGCTGGCTTGCAGTGGTGTGTTGGAGT	505
QY	2773	agtgccgtacgcgagggccalcctccagcaagcgcctggtgtctcttcacagctcaatggttaag	2832
Db	506	GCGAGCCTTCGGGAATCAGACTTTGGAAACAAGCTGTGTTTGGATTAAGGTAAAG	565
QY	2833	agatagcgctgcgtcacctcgtctgttggcagcagcactaagacaaccgcgaagctgcgttc	2892
Db	566	AGCATGGTGACCATTTATCTTATGATATTAACCTTGAGGCTCCAAAGMAAGTCGTTTT	625
QY	2893	cccgagcagctccctcgacagacatacgtcgtctgttgggtctcttgggcccggaggtcacc	2952
Db	626	CCAGAACGTTTACGTGATGACATTCGCACCTCTTGTCAACAGATGCTAGTGATATAGTT	685
QY	2953	accgcgttcacaagatgtgtagcttggccgagcagcctcaagccagcagctgcctctcttc	3012

D	686	TCACGATTTTCACAAGACACACAGAAATGGTTGAGACTCAATATACAGCCTTGCAATTCCTTT	745
Q	3013	ctcagtgaccctctgcccctggttggaaccgggagcttctctcaccgctgctcgggcccac	3070
D	746	CTCAATGATCTGTCTGTGTATGACAGAGAAATTGTTTACCCTATTAAAGTCTGCG	805
Q	3073	tcaagcagggtgccaagcgctccagtcgctcgctccctaatccagcagccctgtgaccctg	313
D	806	TATTAACAGGTCCTTTCAAAGCCTTTACATTTACCATTCAGGATCCGAGTCTTCGTGCTCTTG	865
Q	3133	gcacatgaattcacccgcatcctctgtgacagcaagacatcagtgagaccctccacccccc	319
D	866	AGGCTGGAATTTCTTCGCAATCATCTGCAGTCATAGCAGCATATGTACATTAAACTTACC	925
Q	3193	tctctccccctgtcaccctccagctcgccctccctctctgtgtctctccaccctccag	325
D	926	TCCACCTTACTTACTCCACCTCGCATTCACATCCACCTTCTGTTCTCTCGCATCATCTCAG	985
Q	3253	agctccacctctctccagccaagccccggagcccccaagtlgaccgcatglttcagactgagt	331
D	986	AGTTCTGGATTTTCTACGAATGTCAAAGACCAAAAGATCCAAATATGTTGTGATTATCC	104
Q	3313	ggaccattccggcagcagcagcacttctagcttgggtctctctgtcagcggagctgacatggcc	337
D	1046	GTGCTTTCCGCCCAACAGCATTTATTTGGCAGGACTTGTGTAAACAGGCTGCTGTCATT	1100
Q	3373	ctcgaacatgagtgcttgaaggagcattccgttctgcaagaagaagccatcagtgctgtgac	3433
D	1106	TTAGACCTGTATGCTGGAAGGACTGTTGGATTTCATTAAGAAATCATATAGTACAC	1166
Q	3433	agccctcatgttgccatgaacacagccccgcctacgcgcgaagccaactgtgaagctgct	349
D	1166	AATTACTCTCCAGTCAAGCATCGACCGCGGACTGTGACCTCAAGTAAAGCTGCA	1226
Q	3493	gtgcccagagctgtaccctgccaactgcttcgattgtgcaaggataccttccacagctgcat	3553
D	1226	GTGGCATGTGTATCTACCTCTGATTGGATTATCATGGAACCTGACTCAGCTGAT	1285
Q	3553	gacttctgtgagggcccgaagtcacgcggttcaagaactgtgcctcaatgtcttaccagagaca	3613
D	1286	GATTTTACAGAACTCATCAATCAACAGAGAAAGC-----CAATTGTGTATACCACTGAT	1333
Q	3613	gaagggcgaaggagacattgcgggtlaccataaccctctgtgccaatggtccattgtgt	3672
D	1340	GATTATGAAGTGAAGCGGAGATGATGATGAAGCAACCGTGGCCATCGCATTCGACAGG	1399
Q	3673	ggccccatgacccctgtgctccggggccaagatctcccgaggggccacaaagcctctctgc	3733
D	1400	ACATCGTCCCTCAACTACAAAGCCTGCGCAGTTTCTCTCACGTCAACGAGTGGCAGG	1458
Q	3733	gcaggctgtgccctctctgtcgtgaatcaagccgagacctgtgcgctgtgtgtgtgtgtgt	3792
D	1460	CAACACACTCTACCTTTTCAGCAGAAATCAATCGAAGCCTTTGATCTGCTACTTTGGGTT	1519
Q	3793	ctgaagaacacccgagccgggcctccgagcgctgtgccaatcgacatgaactccccag	3852
D	1520	CTCAAAAATCTCAATGAACAGTTCGTACAGAAAGTGTTTACAGATCTCTCACTCTGAG	1579
Q	3853	ctggagcgtctgttgactgtcgttacccttctccgtgcctctgtgaatcaagaaggagaa	3912
D	1580	CTAAACCGCTATTAGATCTGCTTATCTCTGCTGTCTGTGTTGATATTAAGGAA	1639
Q	3913	aaggccttgaagcatcaacagcctcaactcaaaaaactctgtgataatgaaggcgcg	3972
D	1640	AAAGTGTTTGAACAGATGAATAGCTTGACCTTTAAGAAATCAAAAGACATGAGAGCAAG	1699
Q	3973	ctagaggaagccactgttgataccatcgaggtctgcaagaagaatggttcggcggaagcgt	4032
D	1700	CTTGAAGAGACTATTTCTTGGAGCATAGTGTCCACGCAAGAAATGTGATGGCGAAGCGGA	1759
Q	4033	g-----agaggaccgcttgttgaaatccagaaacgtgtgcctg	4071
D	1760	GGACACCTGAGAGAGCCCATCTCGAAGTGTCTTGGAGATCAAGAAATTTAGAGTGG	1819

4072 cggagagcgtccacacttgaaagcaaacctcagacccgctgtagaagacccaagatgaa 4131
1820 AGGAAAGATATGCTCAGTGGCGTCAAAAACAGAGAACTTGAGAAATCAAGACACAG 1879
4132 atggagacagagccttgtagaaggaacctggaacccagagacccctgtagctg 4191
1880 ATTGAACACGAAACACATGATTATGAAACCTGGCTACAGAAACCAACCTATCTTTTA 1939
4192 gacacatgagatcatcgtgacagcgtgatalgcttcagaagcccgagagacgcttg 4251
1940 GATACATTGAGATTGTTGTTTACAGACCGTTTCTGTACAGGAATCCAAAGAGACATCTT 1999
4252 ggggcaagtctgaaggtgctgctgtaacagccctgggcaagtgcccaagtgccctctg 4311
2000 GGTGGAGTCTAAAGTGTCTACTACACAGCATGGCTGTAAACCAAGTGCAGTTATCTTA 2059
4312 cagcatgagcctgacccagagagggccctgctgctcaagttcccggaagctgctgtag 4371
2060 CAACACTGTTTCTTACACAGAGAGCGCTTGTTCAAAAGTTTCCTGAACTCTTATTTGA 2119
4372 gaagacacgagctgctgtagcagcctgctgtagagctcctcagacatgctgtagcagc 4431
2120 GAGAGACAGAGCAGTGTGCTGATTATGCTCAGGCTTCTCCAGACATGTAGCAGTAGC 2179
4432 atcagacacatccgacgacgacgacgctgctgctgctgctgctgctgctgctgctg 4491
2180 ATCGGTACAAATACGTCACAGCGCACTGCTCTTACCTATTAATGAGGCAAACTTT 2239
4492 gagatcgagcacaacttgccgctgtagaagatgacacatgctcctcctgctgctgctg 4551
2240 GAGATTGGGAAATACCTTTGCGAGGGTTAAATGCAAGTAAACATGTCACATTCCTCTG 2299
4552 gtaggagacgacgacacatcagtagaagacacatgctgtagcctcctcctcctcctc 4611
2200 GTGGGACATCTCAGAAATTTAATGAAGAAATCTTAAGAGCTTCTTAAGACATTAATTG 2259
4612 acctatgctgaagagacagggcgctgtagaagacacacatcctgcaagacgctgtag 4671
2360 ACAATCTCTAAGAAAGATCTGGAATGAGGAAACAACTTCTGTCAGGTCACAGAT 2419
4672 ctgagctcaacatgacatgacatcctgtagcagacagtagaagatgaaagacacagag 4731
2420 CTGCTTTTCAATCTCCTATGATGATCTTCTGATGATCTGAAATGAAGAAACACAGAG 2479
4732 gacctgagatctcactgacacatcgtacagaatgctccgaggtctacacaggtctac 4791
2480 GATCTGAAATCTGATGATCTAATGTAAGAAATTCCTCAAGGCTTACAGACCTCTCCA 2539
4792 gacctgagatctcactgacacatcgtacagaatgctccgaggtctacacaggtctac 4851
2540 GATCTGAAATCTGATGATCTAATGTAAGAAATTCCTCAAGGCTTACAGACCTCTCCA 2599
4852 gctgagacgagcctgtagcagcagcagcagcagcagcagcagcagcagcagcagcag 4911
2600 GCTGAAGCTGACAGTGTCTAGTCCACTCAGACAGCACTTGTGCTGAAATTTGAGCATG 2659
4912 ctgagacgacgacgacacatgctgtagcagcagcagcagcagcagcagcagcagcagc 4971
2660 CTGAGAGACCGGAAATATCTTCTCTGTGGATGTGAACATTTAGAAATTTCTTAAT 2719
4972 gtagtagagagctcgcacatcctcagacacatcctgctgcgcgcgcgcgcgcgcgcgc 5031
2720 GTTTTGAAGAAATCTGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 2779
5032 tccgggaagacatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5091
2780 TCTGGAATAATCTTACAGAGTCAAGAGCTTGTGGATTTACTGGAACAAGACGCTCTTC 2839
5092 ttacacatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5151
2840 TTCTCTATAGGCTGAGATGATGAAGCACTTAATGAAGTTTCAAAAGTACTTATCTTAT 2899

5152 ctggaagccacccgtgactacaagaagctgtagcagcagcagcagcagcagcagcagc 5211
2900 CATGAACCTATTCGGATGCAAGAAACATATCCAACTTATGTAAGTCAAGAAAGA 2959
5212 ttcacagatcatgacacagagctccgctgtagcagcagcagcagcagcagcagcagc 5271
2960 TTGAGCAAAATTTCTTATCAGAGTACGCTGGAGCGAGTGTGGACCTATTTTCTGT 3019
5272 gtaggctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5331
3020 GTTGCTTTTATGGAACCAAGTTCGGGATTTGGATGAACAAGATTTGTTTCAAGAG 3079
5332 ccacatgacagaaagctgtagagatcacaacccgctgtagagagctcacaagcagaga 5391
3080 CCTGCATTAACCAAACTTCAGAGATATCTACAGATTGAGAGGATTTTTCGAGAAAGA 3139
5392 tttagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5451
3140 TTTGGAGAGATGTGGTGAAGTAAAGACCTTAATCCTGTAGACAAAGTGAATTA 3199
5452 gactacaaagcctacatcacagatcagatgtagaagcagcagcagcagcagcagcagc 5511
3200 GATCTTAACAAAGCATATATTCAGATTACCTATGTGAGSCATCTTTGACACATATAG 3259
5512 ctcaagacccggtgtagactctgtagcagcagcagcagcagcagcagcagcagcagc 5571
3260 ATGAAGACAGATACCTATTTGACAAAAATTAACATCTTCTGATATCATGATCTGT 3319
5572 acgcgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5631
3320 ACACCTTTACTTATGAGAGCGCTGCTGAGGAGACTTCAATGAACATTAACAAAGAG 3379
5632 acgctgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 5691
3380 ACATTTCTGACATACCTGATGAGCGCTTCTTATATTAACAAAGGTCATATGCTCAT 3439
5692 cggagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5751
3440 AAAGAAAGATCATCTTAACACCAATTAAGTGTATGAGGATGCAAGAAAGACA 3499
5752 cggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5811
3500 CAGGAGTTGGCACTTTGCAACACATCAGGATCCCGAGACCCCAAAATGCTTACATGTA 3559
5812 ctcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 5871
3560 CTCCAGAGGATCTGTAGGCAACACATGAAATCAGGGGCTTTGGAAGTTGCCAGGTTT 3619
5872 ttagcagagatcccggaagaccccaagctctcgcgcatcacacaatgctgctgctg 5931
3620 CTGCTGAATATCTGATGAGTGAAGCTCTTCAAGACATCATTAATGAGTGTGCTG 3679
5932 ttcagagactctgcaagaatgtagagctgctgtagcagcagcagcagcagcagcagc 5991
3680 TTTAAAGATTATTAATAAGGCTGGAAGTGTCTTAAGAAAAAATGAAGCTTAATTGG 3739
5992 ccgagacagagagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6051
3740 CCGGATCAAAAGAGATTAAGAGGAACTGGAGAGAAATATCATTCGCTTAAGAGGCG 3799
6052 ctgagccctgcttaccagcagcagcagcagcagcagcagcagcagcagcagcagcagc 6086
3800 CTACAGCACTGATCAACAGAAAGATCCCTCAGTT 3834

RESULT 8
AX172880 4144 bp DNA linear PAT 03-JUL-2001
LOCUS AX172880
DEFINITION Sequence 7 from Patent WO0142297.
ACCESSION AX172880
VERSION AX172880.1 GI:14597915
KEYWORDS human.

4192 gaacacttgagatcatcgtgcagacggtgatgtcttcaagaagcccgagagcgcttg 4251
1760 gATACATTAGAGATTGTTGTTTCAGACCGTTTCTGTACGGAAATCAAGAGAGCACTTCTT 1819
4252 ggggacgtgctgaaggtgtgtgtacagcctggacgtgcccagatgtccctctcttg 4311
1820 GGTGGAGTCTMAAGGTCTACTACACAGCATGCCCTGTAAACCAAGTGCAGTTTATCTA 1879
4312 cagcatgtgcctgcacccagagggccctgtgtccaaagttcccggaagctgtgtcgag 4371
1880 CAACACTGTTTTCTACACAGAGAGACCTGTTCGTTCAAGTTTCCGAACTTTATTTGAA 1939
4372 gaggagacagagctgtgtgcgacccgtgtccgtgacctcctacagacactgtgcgcgc 4431
1940 GAAGAGACAGAGCATGTGCTCTGATTATATGCCCTCAGGCTTCTCCGACACTGTAGCTAGC 1999
4432 atcagacacatcgcagacgacgacgacgacgacgacgacgacgacgacgacgacgac 4491
2000 ATCGGTACATATAGTCACACCCAGTGCCTCCCTTACTACTATAGAGCAAAACTTT 2059
4492 gggatgcgcacacttgcctgtgaaagatgcaggtcacacatgtctctcgtccctg 4551
2060 GAATTTGGAAATTAATTTCCAGGGTTAAATGACAGGTACCAATGCTATCTCTCTTG 2119
4552 gtggagacagcgcagacactcagtgaaagacacgtgcagctcactcaaaaacatctc 4611
2120 GTGGGACATCTTACGAATTTTATGAAGAAATCTTAAGACGTTCTCTAAGACTATATG 2179
4612 acctatgtgaagagacatggggctgcggagacacacttcgcagagcaggtccagagac 4671
2180 ACATATGCTGAAMAGATGTGGATTTGAGGGAAMCAACTTTCGTATCAGGTAGCGGAT 2239
4672 ctatgttcaacactgacatgatcctgacgacagcgtagaagtagaagaaacacagag 4731
2240 CTGGTTTCAATTCCTCATATGATTTCTTCTGATACTGTGAAATGAGAAACACCGAGAG 2299
4732 gaccctgagatctcatcgaactcatgatacagaattgcgcccgggtacagaggtcaca 4791
2300 GATCTGGAATATGTTGATTTGATCTAATGTACAGAAATGCCAAGGGTTACAGACACTCTCA 2359
4792 gacctgcgctgacctggtgtgcagaacacggccgggaagacgcggaagcttgggcaacac 4851
2360 GATCTGGATTTGACCTGCTGTGAGAACATGCGCAGGCAAGCAACCAACGAACCAATCAT 2419
4852 gccgaagggcccgacgtgacgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 4911
2420 GCTGAAGCTGCACAGTGTCTAGTCCACTCAGCAGCACTTGTCTGTAATATTGACATG 2479
4912 ctgcagagacacgcgcacactgcgcgcgtgtgcgttctcctccagaacatctcaaac 4971
2480 CTGAGAGACCGGAAATATCTTCCGTGCGATGTGTAACATTTCAAGATATTTCATCTAAT 2539
4972 gtgctagaagaagtcgcacatctccgcagacatccctgtgcgcgcgcgcgcgcgcgcgc 5031
2540 GTTTTAAAGAAATCTGGGCTTCAGATGATGTGTATCTTCAATGAAAGATATCTGTC 2599
5032 tcgcggagacactcactgacgtgcgcgcgtgtgtgaggtgtgtgtgcgcgcgcgcgcgcgc 5091
2600 TCTGGAATAATCTTACTGAGTCAGGACTGTGGGATTTACTGGAACAGAGCATGTGCTCC 2659
5092 ttcaacatgggcgcgcgtcctacgagcggtgaatgaggtctacaagaacctcatcccatc 5151
2660 TTCTCTATGGCTGGCATGTATGAAGCAGTTAATGAAGTAAAGTAACTATTATTCATAT 2719
5152 ctggaagcccgctgactacaaagaacgtgcgcgcgtgtgcagacgaacagcgagagcc 5211
2720 CATATGATTAATCGGGATGCAAGAAACTATCCAAATTAATGTGTAATCTTCAAGAAAGCA 2779
5212 ttcaacgaagatcatgacacagagttccgcgtgcggagcggtgttgcggagcatctccgc 5271
2780 TTCAAGCAAAATTTGTCATCAAGATAGTGGCTGGGAGCGGATGTTTGGCACCTATTTCGT 2839

5272 gtgggtcttccagcgcccaactcgtgtgacctgtgatgagcagaggttgtgtacaagag 5331
2840 GTTGGTTTATATGACCAAGTTCCGGGATTTGGATGAACAAATAATTTGTTTCAAGAGAG 2899
5332 ccatcgatccaggaagcgtgcagagatctcacccgctgtgagaggtctctacaagagaga 5391
2900 CTTGCAATTAACCAACTTTCAGAGATTTCTCACAGATTTGAGGGATTTTACGAGAAAGA 2959
5392 ttgtgcagcagcgtctgtgatatatcaaaagactctaccctgtgcgaagttccacgtt 5451
2960 TTGAGAGAGATGTGTTGAAGTAAATCAAGACTCTTAATCTGTAGACAGTGAATATTA 3019
3020 GATCTTAACCAAGCATATATTCACATTAACCTATGTGAGGCAATCTTTGACATATGAG 3079
5512 ctcaagagacgggtgacctctctgtgacgcgaactatggtgtgcacatctcgtctgc 5571
3080 ATGAAGGACGAATCACCTATTTGACAAAATTTACAAATTTGTCGATTCATGCTACTGT 3139
5572 aagccgttcaacgcgcgaatgggcgcgcacacggaggtgcgcgcgaacaaacgaagcgt 5631
3140 ACACCCCTTACTTTAGATGGCCGTGCCATGGGAACTTCATGAACAAATTCAAAAGGAG 3199
5632 acgctgtcagcaccgcgaacgccttccccaactcaagactcgcacgcgtgtgtgcac 5691
3200 ACATTTGACTACGTCTCATGCTCTTCTTATATTAACCAAGGTTCAATGTACTCAT 3259
5692 cggagagagacggtgtgcagcgcagtgaggtgcgcacatgcagagacatgcacagatg 5751
3260 AAAGAGAGATCATCTTAACCAACCAATTTGAATGCTTTGAGGACATGCAAGAAAGACA 3319
5752 cggagcgtgcgtcttgcacccgcgaaggaaccacagatgtctaagatgtcacagatgtg 5811
3320 CAGGAGTTGGATTTGCAACATCAGGATCCCGACCCCAAAATGCTTCACATGCTA 3379
5812 ctcaaggtctgttagggcccaacgltgaaagaggttcccgaggaggtggccagagttt 5871
3380 CTCCAGGATCTGTAGGACACAGATGATCAAGGGCTTTTGGAAAGTTGCCAGGTTT 3439
5872 ttagcagagatcccggaagaccacagctcttcgcgcacatcaaaaatttgcgtctgc 5931
3440 CTGTTGAAATACCTAAGACCAAGCTCTTTCAGACATCATTAATTAACCTCGACTCTGC 3499
5932 ttcaaggaactctgcagaagaatgtgaggaatgcgtgtgcggaataaagggccgtattgg 5991
3500 TTTTAAAGATTTTAAAGAGGTGTAAGATGCTTTAAGAAAAATTAAGCTTAATTGGG 3559
5992 ccggacccgaagagttacacacggtgagctgcgcgaactaactgcgcgcgtgcggaggt 6051
3560 CCGGTTCAAAAGAGTATCAAAAGGAATTG--GGAAACTATCTTTCGCTTAAAGAGGCC 3617
6052 ctgcagccctgtcttaccagcgctgtcccaagctgatalgacacccccc---accgg 6107
3618 CTACAGCCTTAATCAAGAAAGTCCCTCAAGTATTCACCAAGACAGATTTGTTGCCCTG 3677
6108 cctcaggaactccttgaacagagcaagttccgaaagcgagactctgagc 6158
3678 CCACAGAGATTTCTCAGTGAATGAGCTTTCGCAAAATGATGCTCTAATC 3728

RESULT 9
AX172932 4144 bp DNA linear PAT 03-JUL-2001
LOCUS AX172932 Sequence 59 from Patent WO0142297.
DEFINITION
ACCESSION AX172932
VERSION AX172932.1 GI:14597917
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 4144)

4252 gggagcagctgctgaggctgtgtctgtacagccttggcgagtgcccaagtgccctctctg 4311
1820 ggtggagctgctttaaagctgtactacacacagctgacctgtaacaaagtgccagttaatca 1879
4312 cagcatgagcctgtgcccacagagggccctgtgtccaaagttcccggagctgtgttcgag 4371
1880 caacactgcttttctcacaagagagcccttggttcaaatgttccgaactcttattttgaa 1939
4372 gaagacagagctgtgtgtccagccgtgtccgaagctctctcaagaacactgtgagccgc 4431
1940 gaagaaacagagagagtgctgctgattttatgcctcagaccttctccgacactgtaacagtaac 1999
4432 atcagaacatccgcagcagcagcagcctcgctgtactcgtcgaatgtagcagaacatc 4491
2000 atcgatgacataatggtcacacccagctccctcttacctactaattagggcaaaacttt 2059
4492 gaagatggccagacttcccgctgtgaagatgacaggtacacatgctctctccgctccg 4551
2060 gaagattgggaatgacttgcaggggttaaaatgcaaggtaccaatgtcactatcctccttg 2119
4552 gtgggagcagcagagaaactcagtgaaagacaccttgcaagcttcaactcaaaacactc 4611
2120 gtgggacacatctcagaaatttttaagaagaattcttaagacgttcttaagaactatattg 2179
4612 acctatgctgaagagaaacatgggctcgcggaagacacacttcgacagacagatccagac 4671
2180 acatattgctgaaamaagatctggaattgagggaamaaactttcttgatcaggtccggat 2239
4672 ctatattcaactcagcatgatcctgaacgagcagctgaagaatgaagaatgaagaacacagag 4731
2240 ctggttttcaatctccatattgatttttctgattcgtgtaaaatgaagaaacacacagag 2299
4732 gaacctgaagctcctac 4791
2300 gatcctgaatattgtgattgattgattgattgattgattgattgattgattgattgattgatt 2359
4792 gaacctcgagctgacctgtgtgtgacaaatgagccggaagacagcagagcttggcaaccc 4851
2360 gatctggattgacctggtgtgacaaatgagccggaagacacacacacacacacacacacac 2419
4852 ggcgagagccgagctgacatgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 4911
2420 gctgaagctgacacagtgctgac 2479
4912 ctgagagacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4971
2480 ctggagagacccggaatattcttctggtgagtgatgaaacatttcagaaattttcatctaat 2539
4972 gtgctgaagagagtcgac 5031
2540 gtttttgaagaatcttgcggtctgacatgattgattgatttctcagatgacagagcttattgc 2599
5032 tccggaagacactcactgagctgtgagtgagtgagtgagtgagtgagtgagtgagtgagtg 5091
2600 tctggaataatatttactgagctgagcagcttctggaattactggaacaagacagctgctcc 2659
5092 ttaacatggcgggcgctctcagcgcggtgaatgagtgcttaacaagaacctcaacccac 5151
2660 ttctctatgctgctgacatgattgaagcagtttaagatttcaaaacttactatttctatt 2719
5152 ctggaagccacagctgac 5211
2720 catgaagctaatcgagtgacaaagaaacttccacaattcatgtaactgtcaagaagca 2779
5212 ttaacaaagatcatgacacacagatctccgctgtggagcgcgltctcgagacgtatctccg 5271
2780 tttagcctaaattgttctgacagactgctgagagcgagtggttgacaccttattttccg 2839
5272 gtggagcttcaagcgccacactcggtgacctgtgagagagagatgttggttcaagag 5331
2840 gtttggttttatgaaacaaagcttgcgggatttgatcaaaacaaattttgtttcaagag 2899
5332 ccatcgatcaagagctgagagatctcacaccgctgagagagtgcttcaacgagagag 5391

2900 cctgcaataaacaaacttgacagagatattctcacagattggagggattttacgagaaga 2959
5392 tttagcagcagcgtctgtgagatatacaaaagctcttaccctgtggaagaagctcaagctt 5451
2960 tttagagagagtggtgttggaagtgatcaaaagctcttattctctgagcaagtgaaattta 3019
5452 gactacaagaagcctacatccagatcagatgtgtgaacagcttacttgaatacctacag 5511
3020 gatcttaacaaagcattatattacagattacctatttgagagccattttgacacattatgag 3079
5512 ctcaagagccgggtgacctactttagacgcgaactatggtgcttgcaatctcgttctgc 5571
3080 atgaagcacaatcaccctattttgcacaaaatttacaatcttctcgaattcattgactgt 3139
5572 agccgcttcaagcggatggcgcaacagcgagagctgtcccgcaacacaaacgctaaag 5631
3140 acacccttacttttagatggccgctgcccattggggaacttcatgaaacaaattcaaaagag 3199
5632 agcgtctcaagaaacagcagccttccctacatcaagactgcacatcgltgtgcac 5691
3200 accattctgactagctgctcatgcttcttcttattttaaacaagggctaatgtctactat 3259
5692 cggagagagagcgtgtgtgacgcagtgagggtggccatcgagagacatgcagaagaaga 5751
3260 aaagagagatcatctttacacacatgaaattgcttattgagacatgcagaaagaaaca 3319
5752 cggagagcgtcttccacacagcagacccacacagatgtcgaatgtctacagatgtgtg 5811
3320 cagagattggcattttgcaaacatcagagatcccgacaccccaaaatgctttcagatgcta 3379
5812 ctcaagagctctgtagggccacacagctgaaacaggttccctggagagtggtgccaggtt 5871
3380 ctccagagatctgttagggcaccacacagatgaaatgagggccttttggaagttgcccaggttt 3439
5872 tttagagagatcccggaagacccacagcttcccgacacacacaaatgctgagctgc 5931
3440 ctgctgtaaatctcagtgagacccaagctcttcagacatctatattttaaactgcagctgc 3499
5932 ttaagaagcttctgaagaatgtgagagtgctgtcggaataaataagggccgtattggg 5991
3500 tttaagaattttactttaaagtgatgaaatgctttaaagaaataaagagcttattggg 3559
5992 ccgagccgaagaggtaccacogtgagctgagagcagctctgcgctgcgagagct 6051
3560 cgggttcaaaagcattatcaaaagggaattg--ggaaactatcttccgctttaaagagcc 3617
6052 ctgaccccgcttaccacagcgcctgcccagctgtagtgcacacacccc---accgg 6107
3618 ctacagcccttagatcacagaagttccctcagttatccaaagccagatattgctgtccctg 3677
6108 cctcagaactccttgaacagagcaagtttcogaagagcagactctgagc 6158
3678 ccacagagatctcttcagtgcaatgagcttttcgcaaaatgagatcttcaaac 3728

RESULT 10
AX172980
LOCUS AX172980 4200 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 107 from Patent WO0142297.
ACCESSION AX172980
VERSION AX172980.1 GI:14597958
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequence.
REFERENCE
1 (bases 1 to 4200)
AUTHORS
Lu, P., Garman, J.D. and Candia, A.F.
TITLE
Clasp-3 transmembrane protein
JOURNAL
Patent: WO 0142297-A 107 14-JUN-2001;
Arbor Vita Corporation (US)
FEATURES
location/Qualifiers
1..4200

```

/organism="synthetic construct"
/db_xref="taxon:32630"
/note="polynucleotide fragment"
BASE COUNT      1262 a      844 c      907 g     1187 t
ORIGIN

```

Query Match	22.7%;	Score 1446.6;	DB 6;	Length 4200;
Best Local Similarity	64.5%;	Pred. No. 4e-221;		
Matches 2239; Conservative	0;	Mismatches 1199;	Indels 33;	Gaps 4;

QY	2713	cgcaactcggcagcagaagctgcgtctacagagagcctgctctcgaatgagtggtcgaacgc	2772
Db	321	CGCTTACCAACTTAAAAAGCTTTTTCACGAGAGAGCTGGCTTTCGATGGGGTTGTTTGCACT	380
QY	2773	agtcgcctacgcagagagcactcctcccaagcagcgtgtctctctcccaactcgaatggtgag	2832
Db	381	GCGAGCGTTTGGGAATCAGGTTTGCAMCAAGCGCTGGTCTTTTGTGATTAATGGTAAAG	440
QY	2833	agatagcgctgcgaactcgtctgtcttggccagcgactagacaacaccccgcaagctgcgcttc	2892
Db	441	AGCATGGTGACACATTATTAACCTTAAATGATTAACCTTGAAGGCTCCAAAGAAAGTCGTTT	500
QY	2893	cccgagacgtctccctggagacacactgcgtctggcggtcctggtggcctggaaagtcatc	2952
Db	501	CCAGAACGTTTCAATGGATACATTTGCACTCTTGTCAGACAGATTGCTAGTGAATATAGTT	560
QY	2953	accgcgtccacaaggaatgtggagctgagcgcagacactcaacgcgcagcctgctctctc	3012
Db	561	TCACGATTTCCAGAAAGACACAGAAATGGTTGAGAGACTCAATATCAAGCCTTGCAATTCGTT	620
QY	3013	ctaaaggaactcttgcccttggtagccgggcttctgtctccagcctggtccggcccaac	3072
Db	621	CTCAATGATCTGTGCTGTATATGACACAGGATTTGTTTTAGCGTTATTAATCCTGCG	680
QY	3073	tacaagcaggtggccacgcgcgtctcaagtcgtcccttaatccagcagccctctgacctg	3132
Db	681	TATTAACAGGTGCTTCAAAAGCTTAACTATTCAGATTCACCAAGTGTCTGTGTCCTTG	740
QY	3133	cgcataggaattcaaccgcatcctctgtgcagccagcaactcgtgacccctcaactccc	3192
Db	741	AGGCTGGATTTTCTACGAATCACTGCAAGTATGAGCACTATGTATCACTTAAACTTACC	800
QY	3193	tgtgcacctgtgcaactccagcctccgcctccctctgctctcccaactccac	3252
Db	801	TGCAGTTACTTACTCCACCTGCATCTGCATACCTTGTGTTTCTTCTGCAACATCTCAG	860
QY	3253	agctccacctctccagccaagcccccgagcccccaaggtgacccagcatgttgcactgagt	3312
Db	861	AGTTCTGGATTTTCTACGATGTACAAAGCAAAAGATTCGAATATGTTGAAATTATTC	920
QY	3313	ggacacattccggcagcaggaactcctctagctggtcctcctgcgcgaagggctgagcctg	3372
Db	921	GTGCTCTTCCGCACAAAGCATTAATTTGGCAGGACTTGTTTAAACAGAGCTGGCTGTAT	980
QY	3373	ctcgaacccctgagagctgaaagggcatctcctgttgcacaagaagagccacagtgctgac	3432
Db	981	TTAGACCCCTGATGCTGAAGGACTGTTTGGATTGTCATAAAGATCATCATATATGGTACAC	1040
QY	3433	agcctgcatctggtgcatagacatgtaaccccgctacgcgcgaagggcactgtgaagtcgt	3492
Db	1041	AATTTACTCTCCAGTCCAGACTCAACGCCGGGTACTCTGCACCTCGAATAAAGGCTCGA	1100
QY	3493	gtggccggagctcttactctgcaactgcttctcgattgcacgggataactctgccaagctgcat	3552
Db	1101	GTGGCCATTTTATCTACTCTCTGATTTGATTTATCAATGCAAACTGTACCTCAGCTGTAT	1160
QY	3553	gaacttgctgagagggccaggtcgaagggatgcgaagactgscctaaatgctgtgactcaagaca	3612
Db	1161	GATTTATAGAAATCCACAAATCAACGAGGAAGAC-----CAATTTGTATAGCCACTGAT	1214
QY	3613	gaagcgaaagggacatgvcgggtacatcaaacccctctgtggccatgagccattgctggt	3672

Db	1215	gATTATGAAGATCAGAGCGGAAAGTATGATATAGCAACCGCTTGCCATTGGCAATCCGAGGG	1274
QY	3673	ggcccccctagccctctgctctccgggacgaatctcccaaggggccacaacaggctctctgcg	3732
Db	1275	ACATTCGGTCCCTCAACATTAACAAGGCGCTGGCAGTTTCTCCTCAACGCAAGAGTGGCAGG	1334
QY	3733	gcagcgtctgacctctctctctgaagtaaacgggacctctgagcgltgltgtgtgtgggtg	3792
Db	1335	CAACACACTACTCTTTTCACGACAGATCAATCGAAGCTTTTATCTGTCTACTCTTTGGGTT	1394
QY	3793	ctgaataacaacccgagccggcgctctctgcgacgctgggccaactgacctgcaactcccccag	3852
Db	1395	CTCAAAAATAGCAGATGAACAGACTTCTACAGAGTGGTTTACAGATCTCTCACTTTGGAG	1454
QY	3853	ctggagacgctctctgtgacctgtctacacctctgctcgctcgctcttgagtataaggggaa	3912
Db	1455	CTTAACCGGCTATTAGATCTGGTTTATCTCTGTGTCTTGGTTTGAGATTAAGGGAA	1514
QY	3913	aagccctcttgaacgcatcaacagccctcaatccaaatactctgatatgaagcgcg	3972
Db	1515	AAAGTGTTTGAACGATGAATGATAGCTTTGACCTTTAAGAAATCAAAAGACATGAGCAAG	1574
QY	3973	ctagagaaagccatctctggtatccatcggagctcgacaagaaatggttcggcgaagtcgt	4032
Db	1575	CTTTGAAGAAGCTATTCTTGCGAGCATATGCTGCACGCAAGAAATGATCGCGCAAGCCGA	1634
QY	4033	g-----agggggcccggtttggaaatccagagagaagctgctgcg	4071
Db	1635	GCACAGCTCGAAGAGACCCACTCTGGAATGCTTCCTTTGGAGTCAAGAAATTTGGAGTGG	1694
QY	4072	cggaaagcgctacacactgtaagaaacactatagaacgctgtgaaacaagaccaagatgaa	4131
Db	1695	AGGAAAGATATATGACTACCTGGCGGTCAAAAACACAGAAAGCTTGACAAATCAAGACAGG	1754
QY	4132	atcgaaaccgagcctctgtgtgaagggaacctcggaaccggaggaacctgagttctctg	4191
Db	1755	ATTGAAACAGGAAGCACTGATTTGATGAAACCTGGCTTACAGAAAGCAAACTTAATCTTTTA	1814
QY	4192	gacacactggaatcatcgtgcagacggttgatgcttcagaagcccggaagagcgtcttg	4251
Db	1815	GATACATTAGAGATTGTGTTTTCAGACCGTTTCTGTAAAGGAATCCAAAGAGAGCTTCTT	1874
QY	4251	ggggcagctgtcgaaggttgtgtctgtacagcctctgggcagctgcccagatgtcccctctctg	4311
Db	1875	GGTGAAGTCTMAAAGTCTACTTACACAGCATAGGCTCTTAACCAAAGTGCAGTTTATCTA	1934
QY	4312	cagcactgcacctgcacccacagaggcccttgtctcaagttccocggagctgctgtcgag	4371
Db	1935	CAACACTGTTTGGCTACACAGAGACCCTTGGTTTCAAGTTTCCGAACTCTTATTGAA	1994
QY	4372	gaagaaacggagctgtgtgcgaacctgtgctgtaggtcctcagaacctgtggaagccgc	4431
Db	1995	GAAAGACAGAGCACTGTGCTGATTAATGCGTCAGAGCTTCCGACACTGTAGCAGTAGC	2054
QY	4432	atcaagcaccatccgacgacgacgacggccctgcgtacagccgctcatatcgacgaagcttc	4491
Db	2055	ATCGGTACAATACGCTCACACCCAGTCCCTCTTACCTACTAATGAAGCCAAAACCTT	2114
QY	4492	gagatcgccacaactctgcccgtgtgagaatgcaagctacatgctctctctgctccg	4551
Db	2115	GAGATTTGGGAATTACTTTGCCAGGTTTAAATGCAAGTACAAATGTCACTATCTCTCTG	2174
QY	4552	gttgggacgacgcagaaactcagtgaagaagcacctgcgagcttcaactcaaaacacatctc	4611
Db	2175	GTGGCCACATCTCGAATTTTAATGAAGAAATCTTAAGACGCTCTCAAAACATAATATG	2234
QY	4612	accatgtctggaggagaaatgggctctcggaacgacaccttcgaagcgaggttccagaac	4671
Db	2235	ACATATGCTGAAGAAGATCTGGAATTTGAGGGAAACAAACATTTCTGTGATCGCTCAGGAT	2294
QY	4672	ctgatgttcaacctctgcacatgatcctctcagggacaacggttgaagatgaaagaaacaccagag	4731
Db	2295	CTGGTTTCATCTTCATATATGATCTTTCTATCTGTAAGAAATTAAGAAACACACAGGAG	2354

OY	4732	gaccttgatgagctccalcagccctcatgtataagaattgcgcgggctaccaggctccacgg	4791
Db	2335	GATCTGTGAAGATGTTGATTGATCTTAATGTACATAGAAATTTGCCAAGGTTTACCAGACCTCTCCA	2414
OY	4792	gaccttcgagctgaccttgattgacagaacatggtccgggaagaacagcgaggtctggacaacac	4851
Db	2415	GATCTGGCATGTGACTCGTGTTGCAGAACATGTGGCAGACACTCCAGAACAGAACTCAT	2474
OY	4852	gcgcagagccgcgcccaagtgcattgtgacacgcgcgcgcctcgttgcttgatgaacctgcgcctg	4911
Db	2475	GCTGAAGCTGCACGTTCTCAATGTCACATCCAGCAGCACTTGTTGCTGAAATATTTGGAGCATG	2534
OY	4912	ctcgagagccacgacccacccgcgcgcgcgtggcgcttgcttccctccagaacatccatccaac	4971
Db	2535	CTMGAGGACCGGGAATATCTTCTCTGTGCGATGTGTAACTTTCAGAAATTTTTCATCTCAT	2594
OY	4972	gtctagtagagagtlccgcacatcccgacacatccctgtgcgcgcacgagagaggtctctgc	5031
Db	2595	GTTTATGAAGATCTGCGGCTCTCAGATGATGTTGTTCTCCAGATACGAATGAAGATCTGCG	2654
OY	5032	tcgcgggaagcactctactctgagctcgggagctgtagtggcttccttgaaacagcagccgcctac	5091
Db	2655	TCTGGAATATCTTTACTGTGATGAGCACTTGtGGATTACTGGACACAGCACTCTCTTCC	2714
OY	5092	ttcaccaatgggcgcgcctcctacgcgcgtgtgaatgaagttacaaagaacctcaatcccatc	5151
Db	2715	TTCCTATGCTGCTGCATGTTATGAAGCAGTTAATGAAAGTTTACAAAGTACTTATCTCTATT	2774
OY	5152	cttgaagaccacacgctgtactataaagaagcttgcgcgcgtgtgcaacgcgaactctgagagcc	5211
Db	2775	CATGAAGCTATATCGGGATGTCAGAAAGAACTATCCATTCATATGTAAGAACTTCACAAACCA	2834
OY	5212	ttcaccaagaatcaltgcacacagagttccgcgtctggaagcgcgtgttcgggaacgtattccgc	5271
Db	2835	TTTCAGCAAAATTTGTTCAATCAGAGTACTGCGTGGAGCGGATGTTTGGCACCTATTTTGCT	2894
OY	5272	gttgagcttctaagcgcgcccaactcgcgtgaccttgatagcagagagtttgtataaagag	5331
Db	2895	GTTTGTTTTATGTGAACCAAGTTTCGGGGATTTTGGATGAACAAGAAATTTGTTTCAAGAG	2954
OY	5332	ccatcagtcacagagctcgcgcagagatctccacacgcgcctggaagagttcttcaacgcgagga	5391
Db	2955	CTCTGCATTAACCAACTTTCGACAGATATCTCACAGATTGGAGGATTTTACCGAGAAAGA	3014
OY	5392	tttggcgcagcagctcglttgatgaltatcaagaactcttaccctctgtagacaagtccacgct	5451
Db	3015	TTTTGAGAGAGTATGGTTTGAAGTAAATCAAAAGACTCTAATCTCGTAGACAGAGTAAATTA	3074
OY	5452	gactcacaaaaggcctatacccaagatcacaogtatgtgaaacgcttacttgatccctaagag	5511
Db	3075	GATCTTAACAAAGGATATATTTACGATTACTCATGTGACCACTTGTTCACACATATATAG	3134
OY	5512	cttcaagagcccggtgacactacttgacgcgaactaagtggtcttcgcacactccctgtctgc	5571
Db	3135	ATGAAGAGCAAGAAATCACTTTCGACAAAAATTTACAACTTCGTCTGATCTCATGTACTGT	3194
OY	5572	acgcgcgttcaacgcgcgatggcgcgacacagcgggagctgtcccgagcaacaacaagcgtlaa	5631
Db	3195	ACACCCCTTTACTTATGATGTGCGCTGCCCATTTGGGGAACCTTCATGAACAATTTCAAAGGAC	3254
OY	5632	agcgtgtctagagcaccgcagaccgcctccctccatataagactcgcataccgctgltgtgcac	5691
Db	3255	ACCATTTTGACTACGTCTCATGCTTTTCCCTTAATTTAAACAAAGGCTCAATGTCACTTAT	3314
OY	5692	cggagagacacggtgtcgtacgcgcagctggaagtggtgccaatgcaggaacatgcagaagaagca	5751
Db	3315	AAAGAAAGATCACTTAATCAACCAATTTGAAGTTGCTATTTGAGAGCATGCAGAAAGACA	3374
OY	5752	cggagagctgctccttgcgcacgcgcagacagacccaacagatgctaaagatgctacagatggtg	5811
Db	3375	CAGGATTTGGCATTTTGCATACATCACTCGAGGATCCCGCAGACCCCAAAATCTTCATGATGTGA	3434

FEATURES	SOURCE
<p>Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: http://image.lnl.gov</p> <p>Series: IRAK Plate: 39 Row: 4 Column: 18</p> <p>This clone was selected for full length sequencing because it passed the following selection criteria: similarity but not identity to protein.</p> <p>Location/Qualifiers</p> <p>1..1900</p> <p>/organism="Mus musculus"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="IMAGE:5005183"</p>	<p>Db 3733 CCACAGAGATTCTCTCAGTCGATGATGCTTTCGCAAAATGATCTCTTAAC 3783</p> <p>Qy 6108 cctcaggaactccttgaaacaagagcaagttccgaagaagcaagcctctgagc 6158</p> <p>Db 3673 CTACAGACCCCTTATGATCAGAAAGTCCCTCAGTTATCCAAAGCCAGATGCTTGTCCTCG 3732</p> <p>Qy 6052 ctgcagagccctctctaccacagcgcctgcgccagctgatgagcaacccccc--accgag 6107</p> <p>Db 3673 CTACAGACCCCTTATGATCAGAAAGTCCCTCAGTTATCCAAAGCCAGATGCTTGTCCTCG 3732</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5872 tttagaagaatcccggaagaagcccaagctctctccgagcalcacacaanaattgcgctcgc 5931</p> <p>Db 3495 CTGTCTGGAATATACGTAGTACCCAAAGCTCTTCAGACATCATTAATAAATCTGGAATCTGC 3554</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTGCTTTAAGAAAAATTAAGAGCTTAATTG 3614</p> <p>Qy 5992 ccgagacccaagagaggtaccacccgttagctctgagcgcgaactactgcgcctctgcggaagct 6051</p> <p>Db 3615 CCGGTTCAAAAGAGATCAAGGGAATGG--GGGAACCTATCTTCGGCTTAAAGAGGCC 3672</p> <p>Qy 5932 ttcaaggaactctgcagaagaatgtgagatgcgtctgcggaataaagaagcctgtatgg 5991</p> <p>Db 3555 TTTTAAAGATTATTACGTAAAGGCTGTGAAGTGTG</p>

|||||
Db 1736 ACCAGACGACCAACCCACCTGCGCTGCTTCGAGATTAT----- 1778
Qy 6248 gctgtgaggatgacacacactgtacttggggctggccctctgctgtgtcccatcgt 6307
Db 1779 --AGCGCCCGCAGTGGAGTCTTAGGAGATAGCCCTCTGTCTGCACTGATGAT 1836
Qy 6308 gtgcactgagtcctccctctttaaataatggtttataaagcaaaaaa 6367
Db 1837 TCATAATTTTGTGTTTGTGTTTAAATTAAAGGACTTTATATACAAAAA 1896
Qy 6368 aaaa 6371
Db 1897 AAAA 1900
RESULT 12
AK055905 3327 bp mRNA linear PRI 31-OCT-2001
LOCUS Homo sapiens CDNA FLJ31343 fis, clone MESAN100101, weakly similar
DEFINITION to Rat trg gene product.
ACCESSION AK055905
VERSION AK055905.1 GI:16550749
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens normal mesangial cells (NHMC56046-2) CDNA to mRNA,
clone lib:MESAN1 clone:MESAN100101.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Ito,K.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Iseno,Y.,
Kawai-Hiro,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuna,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuno,Y., Nagai,K. and Isogai,T.
TITLE NEDO human cdna sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3327)
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
JOURNAL Direct Submission
SUBMITTED (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-1-3951, Fax:81-438-52-3952)
NEDO human cdna sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; CDNA full insert sequencing:
Research Association for Biotechnology (RAB); CDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Eletechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
Source location/Qualifiers
1..3327
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MESAN100101"
/cell_type="normal mesangial cells (NHMC56046-2)"
/clone_lib="MESAN1"
/note="cloning vector: pME18SFL3-primary culture, normal
mesangial cells"
54..2279
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB71042.1"
/db_xref="GI:16550750"
/translation="MEFLSYPRQOHYIAGLVITELAVIILDPABGLFGHKKVIMV
HNLISHDSPRYSDPQIKARVAMLYPLIGITMETVPLDITETHNQRICLAT
DYESGSMISQTVAMAIAGTSVPLTRGSFLITSTSGRHTTSAESSRLLCL
LWYLNADFTVLQWFTDLSVQLINRLDILLCVSGFEKKGKVERNNLSLFFKSK
DMRAKLEALITLISGIAQGEWVRSRQLESPSGSAFSGQENLRMRKDMTHMRNTEK

LKRSAPLIEHIALIDGNLATEANLIIDLTLEYVQTVSTESKESITGLVYLHSM
ACNOSAVYLQHCHEPATRALVSKPELLFEFEQCDLRLRHSSTIGTRSHAS
ASLYILMRONFEIGNNFARVMQVWMSLSLSTSONFNEEFLRSKTLITVAEDL
LEBETPEOVODLVFNHMLISDYEKMEHEDPEMLILMYRIAKGQVSPDLRLT
WIONMAGKHSERNHNAEOCVHSAALVAEYLSMEDKYLPLVCVFQONISVYLE
ESAVSPDVVSPDEGICSGFPRESGLVILBOAAAFSMAGMEVAVYVYLPIH
EANRDKRKLSTIHKGLQEAESKTVHODGKRMKCTRYNCFPTKRGDLDFOEFVKEP
AIIKRLAEISRLEGFYGERPGEVDVYVIXDSNPDCKLDIPNTHIFRLPMWSHTLTH
R"

BASE COUNT 1046 a 640 c 711 g 930 t
ORIGIN
Query Match 19.28; Score 1225.4; DB 9; Length 3327;
Best Local Similarity 65.48; Pred. No. 7.7e-186;
Matches 1873; Conservative 0; Mismatches 956; Indels 34; Gaps 4;

Qy 3245 cctcccaagctccacccctctccagccagcccccagagctgaccagcatcg 3304
Db 1 CATCTCAGAGTCTGGATTCTTCTACGAATGTACAGACCAAAAGATTGCCAATATGTTTG 60
Qy 3305 aactgagtgacccatcccgagcagcactcctagctgggtctctgctgaagagctgg 3364
Db 61 AATTATCGGTGCTTTCCGCCAACAGCATATTGGCAGCACTGTGTAAACAGAGCTGG 120
Qy 3365 cactggccctcgaaacctgaagctgaagggcattccctgctgcaagaagagcctagtg 3424
Db 121 CTGTCAATTTAGACCCCTATGCTGAAGGACTGTTGGATTTCATATAGAAATCATATA 180
Qy 3425 ctgtgacagcctgtctatgtggtccatgacatgacccccctgctgacggagccactgta 3484
Db 181 TGTATACAAATTTACTCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTAA 240
Qy 3485 aggtctgtgtggtcgagctgaactgtccactgtcttgatgacagggatccttgccac 3544
Db 241 AGCTCGAGTGGCCATGTTGTATCTACTCTGATGTATATATCATGGAACCTGACTC 300
Qy 3545 gctctgacatgacttgcctgaagggccaggtcagcggtcaagactgagcctcaatgcttgact 3604
Db 301 ACCTGTATGATTTTACAGAACTCAATCAACGAGAAAGC-----CAATTTGTATAG 354
Qy 3605 cagacacagaagcggaagggacatgcygylaccatcaacccctctgtgacatgacca 3664
Db 355 CCACGATGATATTATGAAAGTGAAGCGAGATGATTAAGCCAGACCGTTCATGAGCA 414
Qy 3665 ttgctgtgtggtcccttagcctctgctcccggtcgacgatctcccaaggggacacaaagg 3724
Db 415 TCGCAGGGACATCGGCTCCCTCACTAACAGAGCGTGGCAGATTCTCTCTCACTGCAACGA 474
Qy 3725 ctcttcgagcaggtgtgctcctctgtgctgaagcagcctgtggtgctgtgctg 3784
Db 475 GTGGCAGGCAACACTACTCTTTTACGCAAGATTCAGAGCTTTGATCTGTCTAC 534
Qy 3785 tctgtgtgtgtaaaacacacagcgcggtctctgagcgtgtggtgacactgacac 3844
Db 535 TTTGGTTCCTCAAAAATGAGATGAAGACGTTCTACAGAAAGGTTTACAGATCTCTCAG 594
Qy 3845 tccccagctggagagctgtgtggtgactgtgtgaacttgcctgctgctgctgtgactga 3904
Db 595 TCTTTGAGCTAAACCGGCTATTAGATCTGCTTATCTGTGTGTGCTGCTTTGAGTATA 654
Qy 3905 aggggaaagagccttgaacgcatacaagcctcacatcacaataatctctgataga 3964
Db 655 AAGGAAAAAAGTGTGTAACGAATGAAGCTTGACCTTTAAGAAATCAAAAAGACATGA 714
Qy 3965 aggcgcgctgagagaaagcattctgtgtacatcagcagagctgcagaagaatggttggc 4024
Db 715 GAGCAAAAGCTTGAAGAGCTTATCTTGAGCATAGTGTCCAGGCAAGAAAGGATGACG 774
Qy 4025 gaagctgtgagag-----agccgctgtggaaatccagacagc 4063
Db 775 GAAGCCGAGACAGCTCAGAGAGAGCCCATCTGGAAGTGTCTTTGAGAGTCAAGAAAT 834

0Y	4054	tgcgcctcgggaagacgcgtccacacacttgaagcaaacctccagcccgctgtgacaaaccca	41220
Db	835	tgagcttgaggaaggaatgatatactcactgcggcttataaaacacaaagaaagcttgcaaatataa	894
0Y	4124	aggaatgaatbgaacacgaagcctctgttgaagaaggaaaccttgcaacacgaagcaaacctag	41830
Db	895	gagcagagattgaacacgaagcactgattgatgaaacacttgccctacagaaacgaacctaa	954
0Y	4184	tggtctcggaaacacctbgaatcaatcgttgaaacggttgatgcttcacgaagcccggaag	42330
Db	955	tcatttttttgatgatactttgagaaattggtttgtagaccccttttctgtaacggatccaaagaa	10144
0Y	4244	gcgcctctgggggcagtgctbgaagtgctgtgctgtacaccccttgggcagtgccceagtgccc	43030
Db	1015	gcattcttgctggcgagctctaaagctcctatcacacagcattggccctgtatccaaattgcag	10744
0Y	4304	tcctcttgcaagcatgagccttgccaccacaaaggccctctgtccaaagtctccggagctgc	43650
Db	1075	ttttatttcacaaactctgtttttgcttaccacagaaagaccttggttttcaaaagtctctgacct	11344
0Y	4364	tgctcgaaggagacaaggaactgtgtgcgaacctgtgcctgaagcgtcctcaagaacctgtg	44230
Db	1135	tattttgaagaagaaacacagacactgctgctgatttatgcttcagcgttctccgacacctgta	11844
0Y	4424	ggaacgcatacagaacacatccgcagcagcaacgagccctgcgtgtatcaacctgcctcagagac	44830
Db	1195	gcagtagcattcggatgacaaatmggttcacacagccacgtgcctcctttaccactaaatgaagcc	12544
0Y	4484	agaacttcgagatcgtgcccacaactttgcgcggtgtgaagatgcaggtcacacatgctctct	45430
Db	1255	aaaactttggaattgggaattgggaatttggccaggggtttaaaaatggcgaatmacaaatctcatt	13144
0Y	4544	cgctccctggctgggaagcagcaggaactctgaatgaagaagcacttgcagacttcaatccaa	46030
Db	1315	ccctccttggtggcgcacatctcagaaattttatgaagaatttttaagagctttcttaagaa	13744
0Y	4604	ccatctccactatgctcgaaagagacatbvggcctgcgcgaagcaaccttcgaagacag	46630
Db	1375	cttatattgacattatgcgaagaagatcttggaatttgagggaacaaacattttcctgtatctgg	14344
0Y	4664	tcaggaagactbgtltcaacctgcaatgactccctgcagcaacacgctbgaagatbgaagagac	47230
Db	1435	ttccagagatctggttttcaattctccatattgatattcttctgatpactggaataatgaaggaac	14944
0Y	4724	accagaggagacctggaatctcatctgacctatbgaagaaatltgcggggctaccag	47830
Db	1495	acccagagagatctcgaataatgttgattgattatattgacagaattgccaaaggtttaccgaa	15544
0Y	4784	gtctaccggagacttcgctgtacctgtgtgagaacatbvgccgggaagcaacgagagactg	48430
Db	1555	ccctccagatctcgcactgactacctcgtggtgcagaaatgcgaagcaagcactcagaaacaa	16144
0Y	4844	gcaaacacgcgcgagcgcccaatgtaatgtagtgcagcgcgcgcgcctctgtagttagtacc	49030
Db	1615	gcattatcctcgaagactgcacagctgtctatctcactcacacacactttgttctgctaat	16744
0Y	4904	tcgcctctgcgaggaacacagcgcacactgcgcgctgtggtctgtcttctccagaacact	49630
Db	1675	tgagactgctcgaagacccggaataattcttctctgsgattgttaacatttcagattatt	17344
0Y	4964	catcaacatgctcgaagagatcgcgcactctccgaacaaactctgtcgcgcgcgcagagag	50230
Db	1735	catcttaattgtyttagaagaattctccggtctcaaatatgatgtggatctccagatgaagag	17944
0Y	5024	gtctctgtcccggaagacacttcaactbagaactbvggctgtgtaggtgtcgtgaacagagag	50830
Db	1795	gtattctgctcgaataattactttactgagacagacttttgsgaattacggaacaagcag	18544
0Y	5084	ccgctactatcaacatbvgcgcggtctcaagagcgttgataaggtctacaagaacctcca	51430
Db	1855	ctgcttctcttctgattctgctgagcatgataaacaacttaataaagatttaacaaagatttta	19144
0Y	5144	tcaccatctcggaaacccacactgtaacaaagactbvgcgcggtgtcgcgcgaacactgc	52030

Db	1915	TTCTCATTCATGAAGCTAATCGGAGATGCAAAATACTATCCACATTCATGTAACCTTC	1974
QY	5204	agagagccttcacacagatcaltgacacagagctccgcgtggagcgcgtgtlctggggacgt	5263
Db	1975	AAGAAGCATTCACGAAATATGTTTCATCAG-----GATGTAAACGGATGTTGGCACCT	2028
QY	5264	attccgcgtgggtccttcacgscgccctctggtagacttgatgagcaagattgtgt	5323
Db	2029	ATTTCGCTGTGGTGTGTTTATGAGAACCAAGTTCGGGGATTTTGATGACAAAGAAATTTGCTT	2088
QY	5324	acaaagagccatcgatcaccgaagcttgcagagatctccacacggcttggaggagtlctaca	5383
Db	2089	ACAAGGAGCCCTGCAGATTAACCAAACTTGGCGAGATATCTCACAGATGTGAGAGGATTTTACG	2148
QY	5384	cggagagatttggagagagcgtcgtgttgagattatcaaaagactctaccctgtggacaagt	5443
Db	2149	GAGAAAGATTGGAGAGAGATGTGTTGAAGTATTAAGAAGCTCTAATCCTGTAGACAACT	2208
QY	5444	ccaagcttgatccaaagaagcctacatccagatcaatcglatgtggaacgttacttgata	5503
Db	2209	GTAATATTAGA-TGCTTAACAGCGATATATATCAATTAATCCTATGTGAGCCATATCTTTGACA	2267
QY	5504	cctcagagctcacaagacccgggtgaccttacttggaccgaactatgggcttgcacattcc	5563
Db	2268	CATATGAGATGAAGAGACAGATCACTATTTGACAAAAATTAACAAATCTTCGTGATTTCA	2327
QY	5564	tgctctgcagccggtltaacgcggatgggcgcgcacacaggggagcttgcgcgagcaacaa	5623
Db	2328	TGTACTGTACACCCCTTTACTTATGATGGCCGTCGCCATGGGGAACTTCATGAACCAATTCa	2387
QY	5624	agcclaagacgccttgcacagccgacacacgcctccctccatcataagaactcgcatacgtg	5683
Db	2388	AAAGGAAGACCATCTGTACTACGCTCAAGCCTTCTTATATTAACAAAGGGTCAATG	2447
QY	5684	tgtcacacccggggagagacggtgtctgacgcgaatggaggtggccaatggagagatigcaga	5743
Db	2448	TCACTCATTAAGAGAGATGATCATTTTAACACCAATTTGAAGTTGCTATTGTAGGACATGCACA	2507
QY	5744	agaagaacaggagagcttgccttctgcacaccgagacagagccacacagatgtctaagtctac	5803
Db	2508	AAAAGACACAGAGATTGGCATTTTGCACACATCAAGATCCCGCAGACCCCAAAATATCTTC	2567
QY	5804	agatgggtcttgagcctctgttgaagccacccgttgaaacaaaggtcccttgaggatggccc	5863
Db	2568	AGATGGTACTCCAGGATCTGTAAGCAGCACACAGTAAATCAAGGGCCTTTTGGAAAGTTGCC	2627
QY	5864	aggtgttlltagcagagatccgcgaagaccaccaagctctctccgcgcatcacacaatgtgc	5923
Db	2628	AGGTTTTTTCGTGTGAAATACCTAGTAGACCCCAACCTCTTTCAGCATCATTAATTAATCTGC	2687
QY	5924	ggcctctctcaagagactcttgaagaanaatgtgagagatggcgttgcggaanaataagagccc	5983
Db	2688	GACTCTGCTTTAAAGATTTTACTTAAAGAGGTGTGAAGATGCTTTAAGAAAATTAATAGCT	2747
QY	5984	tgattggccgagaccagaagagatccacacgttgagcttggagcgcaactactcgcgcgtgc	6043
Db	2748	TAAATTGGGCGGATCAAAAGAGATATCAAAAGGGAATCGAAGAAACTATCATCGCCTTA	2807
QY	6044	gggagagcttgcgaagcccttcttaaccagcgcgtcccccagct	6086
Db	2808	AAGAGCCCTTACAGCCAGTATCATACAGAAAGATCCCTCAGTT	2850
RESULT 13			
AX172978	AX172978	3253 bp	DNA linear
LOCUS	sequence 105 from Patent WO012297.		PAT 03-JUL-2001
DEFINITION	AX172978		
ACCESSION	AX172978.1	GI:14597957	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
	synthetic construct.		
	synthetic construct		

REFERENCE	artificial sequence. 1 (bases 1 to 3253)
AUTHORS	Lu, P., Garman, J. D. and Candia, A. F.
TITLE	Clasp-3 transmembrane protein
JOURNAL	Patent: WO 0142297-A 105 14-JUN -2001; Arbor Vita Corporation (US)
FEATURES	Location/Qualifiers
Source	1..3253
BASE COUNT	1009 a 633 c 705 g 906 t
ORIGIN	/organism="synthetic construct" /db_xref="taxon:32630" /note="Polynucleotide fragment"
Query Match	18.9%; Score 1207; DB 6; Length 3253;
Best Local Similarity	65.2%; Pred. No. 6,6e-183;
Matches 1853; Conservative	0; Mismatches 955; Indels 33; Gaps 4;
Qy	3343 gggctccgtctgacggagctggtgcactggtccctcgcaactcgtgagctgaagggacattctcg 3402
Db	4 GGACTTGtGTtTTCAGAGCTGGCTGTTCATTTTAAAGCCGAGAGCTGAAGAGCTGTTTGA 63
Qy	3403 ttgcaacaagaagcactgactgctgtctgtgcaagccgtcgtatgtggtccatgacactgacccc 3462
Db	64 TTGCATTAAGAAGTATCATATATGGTACCAATTTACTCTCCAGTCCAGCACCTCAGACCCG 123
Qy	3463 cgcatacgccgaagcactgtgaagctcgtgtgtgcccagctgttaactgccaactgcttcg 3522
Db	124 CGGTACTGTGACCTTCACATTAAGAGGCTCCAGTGGCCATGTTGATCTTACTCTGATTGCT 183
Qy	3523 attgcaaggatgacactgtggcagcgtgcgtgactgttgcgtgaagggcccgatgacgttca 3582
Db	184 ATTATATCATGGAACGTGTACTTCAGCTGTATGATTTTACGAAGAACTCACATCAGCAGGA 243
Qy	3583 agactggtcctcaatgctttgactcagaacacaagaagcgcaaggggacaatgcyggtaccac 3642
Db	244 AGAC-----CAATTGTGTATAGCCACTGTATGATTAAGAAAGTAGAGCGGAAGTATGATA 297
Qy	3643 aaccctctgtgtgcatgacgtcgtgtgtgtgccccttgcgccctgtgctcccgggccagc 3702
Db	298 AACCGAGACCGTTGCAATGGCAATCGAGGGAATCGTCCCTCACTAACAAGCCTGGC 357
Qy	3703 atctccagaagggcacaacacagcgtcttcgcgaagcgtgtgacctctgtgtaagtaaacg 3762
Db	358 AGTTTCCCTCTCCACGTCAACGAGTGGCAGGACAACACTACCTTTTCACAGAAATCAAG 417
Qy	3763 cggacctgtgtgctgt 3822
Db	418 CGAAGCGCTTTTGATGTGTACTTTGGGTTCCTCAAAAAGACGAGATMAACAGTTCTACAG 477
Qy	3823 cgtgtggccactgtacctgtgacatccccccagcttggagcgtctgtgtgactgtgtacctt 3882
Db	478 AAGTGGTTTACATATTTCTCAGTCTTGCAAGCTTAACCGGCTATTAAATCTGCTTATCTC 537
Qy	3883 tgcctgtgctctcttltgagtaacaaggggaaaaagcctlttgaacgcatacaacgctcaca 3942
Db	538 TGGTGTCTTGTGCTTTGAGTATTAAGGAAAAAAGGTGTTGAACGATGATTAAGTACTTGAC 597
Qy	3943 ttcaaaaaatctctgtatattgaagcgcggtctagaagaagccatttgggtaccactcga 4002
Db	598 TTTTAAATAATTCANAAACATGAGAGGAAAGCTTTGAAGAGAGCTATTTCTTGAGACATTAAG 657
Qy	4003 gctcgcaagaagaatgttgcgcgaagtcgt-----agagagc 4041
Db	658 GCCAGGCAAGAATGGTATAGCGGCAAGCCGAGACAGCTGCAGAGAGAGCCCATCTGTGAAG 717
Qy	4042 cggtttgggaattccagaagaagctgtgcttgcggaagaagcgttaccacacttgaagcaaac 4101
Db	718 GCGTTTGGAGAGTCAAGAAAAATTTGAGAGTGGAGGAAGATATGACTCAGTGGCTCAAAAC 777
Qy	4102 tcagacgcgctgtgacagagccaaggaatgaatgaacacagagccttgtgtggaaggaac 4161

Db	778	ACAAGAAGCTTGACAAATCACAAGACAGAGATTGAACACGAAGCACTGATTGATGGAAC	837
Oy	4162	ctggcaacccaggaagcaacccagtcgcttcgtgacacacctggatcatcgtgcagacggtg	4221b
Db	838	CTGGCTACAGAGACAACCTATATCTTTTAGATACATTAGAGAGTTGTGTTACAGCCGT	897
Oy	4222	atgcttcaaggaaccccggaagcgtctctggggcagtcgtgaagttgtgtctacagc	4281b
Db	898	TCTGTAAAGGAATCCAAAGAGAGACATCTGTGGAGTGCTAAAAATGCTACTACACAGC	957
Oy	4282	ctgggcagtcgccagagtcgccctctctctggcagcatcgtgcctggcccaagggccctt	4341b
Db	958	ATGGCTCTGTAAACAAATGTGACGTTATCTAACAACTGTTTGTCTACAGAGACCTTG	1017b
Oy	4342	gtgtccaaagttcccgagccgctgcgtcttcggaaagagacagagcgtgtgtccagctgtgc	4401b
Db	1018	GTTCGAAAGTTTCTCTGAACCTTATTTTGAAGAAGACACAGACAGTGTCATTTATGC	1077b
Oy	4402	ctgagccctcctacagacacctgtgacgacgcgatgaacatccgcagcagccagccgc	4461b
Db	1078	CTCAGGCTTCCTCGACACTGATGAGATGATCGGTGATACATAAGGTCACACCCAGTGCC	1137b
Oy	4462	tcgctgtacctgcatactgcagacagactctgagatccggccacaacttctccgtgtgaag	4522b
Db	1138	TCCCTTTACCTACCTAATGAGGCAAAACTTTGAATTTGGGATTAACCTTCCAGGGGTAA	1197b
Oy	4522	atgcaggtccacactgtctctcgtccctgtgtgggagcagcgagaaactcgaatgaagag	4581b
Db	1198	ATGCAAGTACCAATGTCACTATCTCTGTGTGGGCATCTCAGAACTTTATGAAAGA	1257b
Oy	4582	caacctgcagcttcaactaaaacatccctacatctgtcgtgaagagacatctgggctcgg	4641b
Db	1258	TTCCTTAACAGCTTCCTCTAATAAGATATATGCAATATGCTGAAGATGCTCGAATTGAG	1317b
Oy	4642	gacagacccttcgcagagcaggtccagaccctgatcttcaactgcacatgatccctgaag	4701b
Db	1318	GAACAACATTTCTCTGATCAGACAGTCCAGAGATCTGGTTTCAATCCATATGATTTCTT	1377b
Oy	4702	gacacggtgaaagtgaagaagacacacagagagacccttgatgtctcatcactcatgtac	4761b
Db	1378	GATACTGTGAAATGAGGAACACCAGAGAGATCTGAATGTTGATGTATCTAATGTAC	1437b
Oy	4762	agaattgcccggggtctacagaggctcaccgagaccttcgctctgacccgtgtgtcgaacatg	4821b
Db	1438	AGAATTGGCAAGGGTTACCAACACTCTCCAGATCGGATTTGACCTGGTTGCGAGACATG	1497b
Oy	4822	gccgggaagcaogcgagctgtggcaaccaocgcgagggccgcatgtgtgtgcagcg	4881b
Db	1498	GCAGCAAGCACTCGAAGCAGAACATCATCTGTAACCTCACAGTGTATGTGCTCACTGA	1557b
Oy	4882	gccgcctcgtgtcgtgagttactcgcctcgtctgaaagacagcgccacactgcccctgggc	4941b
Db	1558	GCAGCACTTTGTGCTGAATATTTAGACATGCTGGAGGACCGGAAATATCTCTCTGGGA	1617b
Oy	4942	tgcgcttccctccagaaatctccacatccaaogtgtlaagaggtccgcacatccctcagac	5001b
Db	1618	TGTGTAACTTTACAGAAATATTCTCATATGTTTAAAGAATCTGGGGTCTCAGATGAT	1677b
Oy	5002	atccctgtccgccagcagagaggctctgtcccggaagcaacttcaactgagctggggctg	5061b
Db	1678	GTGTATCTCCAGATGAGAGAGATCTGCTCTGGAAAATCACTTTACTGTAGTCAGACTT	1737b
Oy	5062	gtcaggtgtgtggaaacagcgagccggtcacttccacatctggcggggtcctcagagcggtg	5121b
Db	1738	GTGGGATTAATCTGGAACAAGCAGCTGCTCTCTCTATAGCTGGCATGTATGAACAGTT	1797b
Oy	5122	aatggagcttacaagaacctatcccatccctctggaaagcccaacgctgtgactacaagaagctg	5181b
Db	1798	AATGAAGTTTAAAGTACTTATTCCTATTTATGTAACCTAATGGGATGCAAGAAACTA	1857b
Oy	5182	gccgcggtgtcacgcaaacctgcaggaagccttccaagaatcatgcagcagagttccgcg	5241b
Db	1858	TCCACATTTCAATGTAATCTTCAGGAAGCATTCAGCAAAATTTGTTCACTAGCATCTGGC	1917b

QY	5242	tggagagcgctglttcggagacgtaltctccgcgttggtcctctacggcgcgcacctcgtgcac	5301
Db	1918	tggagagcgagatggtttggcaccctatttttcggttggttttatggacaacagttgggggat	1977
QY	5302	ctggaatgaagcagaggttctgtgtacaaggagccatcgatcaacgaagctgcgagagatctca	5361
Db	1978	ttggatatacacaagattttgtttacmagagaccctgcataacccaacttggcagagatattct	2037
QY	5362	caccgagcttggaggagttcttcaacggagagatattbgcgacagcgtcgtttgatatatcaaa	5421
Db	2038	cacagatttggagggattttaccggagaaagatttggagagagatgttggttgaaatratcaaa	2097
QY	5422	gactcttaccctctgtgacaagtcacaaagcttgaactacaaaaggccttaccatccagatcacg	5481
Db	2098	gactcttaattcctctgagcaaaagtgttaatttgatccctaacaaagccattatttcagatttacc	2157
QY	5482	tactgtgaacccgttactcttgaacttaactaagagcctaaggagccgggtacctctcttgacgc	5541
Db	2158	tatgtggagcccatcttttgacacatratgacatagagacagaaatcaccttattttcgacaaa	2217
QY	5542	aactatgggcttcgcaaatctctgtcttcgacgcgcttcaacgcggaatggcgcgacac	5601
Db	2218	aaattacaattcttgctgatttcattgactgatacccttttacttttgaatggccctggccat	2277
QY	5602	ggggagctgcgcgagacaacaacagcgtlaagacgctgcacgcgacccaacgcctccccc	5661
Db	2278	ggggaaacttcattatacaaatttcaaaagaaagaaacccattctgactcagcttcattgaccttgcct	2337
QY	5662	tacatcaagagctcgcatccgtgtgtgcacccgggagagagccgttgttgaagccagtgagag	5721
Db	2338	tatatattaaaacagggctcatatgtcactcatataagaaagagatcatcttttaacaccaattgaa	2397
QY	5722	gtggccatcgagagacatgcagaagaagacacgggagcttgccttttgccacggacggagac	5781
Db	2398	gtttgctatttgagacacatgcagaaaagacacacagagttggcattttgcacaacatcagam	2457
QY	5782	ccacagatagctaaagttgttacaagatggtgtctcaaggtcgtgtaaggcccaacgtgaac	5841
Db	2458	ccgcgacaccccaaaatgcttgcagatgagatccacagggatctgtagggcacacagtgaa	2517
QY	5842	cagggtccccccttgagagtgggcccgaggttttttagcagagatctccgaaagaaacccaagctc	5901
Db	2518	caggggccccttgggaattttcccaaggtttttctgtctgaaattaccttgaatcccaaaagctc	2577
QY	5902	tccgagcatcacacaatattcggtcgtctgtctcaagagacttctgcagaaatgttgaagat	5961
Db	2578	tttcagacatcatatataaactgcgactctgcttttaaagattttacttataaagtgtagaagt	2637
QY	5962	gggcgtgcgaaaaataaaggccctgatttggccggagaccagaagagatcacaccgtgaagctg	6021
Db	2638	gcccttaagaaaaataaagagcttaatttggccgggttcaaaaagagatataaagggaatttg	2697
QY	6022	gagcgcaacttactgcgcgcttcggagagctctgcagccctcttaccagagcctgccc	6081
Db	2698	--gggaactatctttgcctttaaagagcccttacagcccttagatatacagaanaagtccttca	2755
QY	6082	cagctgatatgcacccaacccc---accggccttcaagaaactcgttgaacagagcaagttt	6137
Db	2756	gtttattccaagccagatttgctgtgtcccgccgacacagagatttcccttcagtgcgaatagcctt	2815
QY	6138	ccgaaagcagagccctcgtgac	6158
Db	2816	ttcgcaaaatgagattctataac	2836

SOURCE	Homo sapiens adult spleen cDNA to mRNA, clone:as00026.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 4577)
AUTHORS	Chara,O., Nagase,T., Kikuno,R. and Okumura,K.
TITLE	The nucleotide sequence of a long cDNA clone isolated from human spleen
JOURNAL	Published Only in Database (2000) In press
REFERENCE	2 (bases 1 to 4577)
AUTHORS	Chara,O., Nagase,T., Kikuno,R. and Okumura,K.
TITLE	Direct Submission
JOURNAL	Submitted (24-AUG-2000) Osamu Chara, Kazusa DNA Research Institute Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba 252-0812, Japan (E-mail:cdna1fo@kazusa.or.jp, URL:htp://www.kazusa.or.jp/NEDO, Tel:81-438-52-3913, Fax:81-438-52-3914)
COMMENT	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'-, 8' 3'-end one pass sequencing and clone selection. Kazusa DNA Research Institute.
FEATURES	Location/Qualifiers
source	1..4577

	BASE COUNT	FULLY ALIGNED
ORIGIN	1297 a	1037 c 1077 g 1166 t

	Query Match	18.3%	Score 1167.8;	DB: 9;	Length 4577;
	Best Local Similarity	60.8%;	Pred. No. 1.le-156;		
	Matches 2055; Conservative	0;	Mismatches 126;	Indels 57;	Gaps 7;
QY	2724	cagcagaactgcctccacgagagcctgctcctgaatgggtgtgttcacgacagcagtgcctgatac	2783		
Dd	112	CAAAAAGCATTTCCATGTGAGGACCTTGCCCCCTCAGATGGTGTGCACGCCGAATTGGTAG	171		
OY	2784	cgaagccatctctccacgacgacctgattcttctccagctcctgtgtgaagatgatcgacct	2843		

Db 172 AGAAGACGCTCTTACAGTATGCGCTGTTCTCTTGAAGCTCTGGGAAACATGCGCCA 231
OY 2844 gcaactgtgcttggccagcgactatagacacaccccgcaagcttgcttccccgagcgt 2903
Db 232 GCACGTACATTAACATGACAAACGGGACAGTTTTCGGAGSACTGGTTTCTGACCGGTT 291
OY 2904 ccttgagacacatctgcttctgtgtgctctgtgtggtcgtgaggtatcaccgtgt-- 2960
Db 292 CATGGATGACATTAATCTATTGTTAATGTGTGTCACCTCGGAAATTCGACGCTTTTAGT 351
OY 2961 -----ccacaagatgtgtagcgtggccgagcacctcaacgagcagcgttggcttctcct 3014
Db 352 AAAACCAAGAAAGAAATGAACAGCGGAAATGAAATCAATCAGCTTGGCTTCTCTT 411
OY 3015 cagtgaaccttcttcccttgtgtggaacggcggtcttgcttccacgttgctcgccgaccta 3074
Db 412 GTATGACCTTCTTCCCTCATGATCGGGGCTTGTGTTAACCTCATAGACATTTATG 471
OY 3075 caaggaagtgccagcgcgctccagtgctgctccctaalcacagcagcctgtgtaccctg 3134
Db 472 CAGCGACGCTGACGCAAGCT-----CAGTAACCTTCCAAAGCTCATTTCCATGAG 522
OY 3135 catggaattcacccgcatcctgtgtgacccagcagcactacgtgacccctcaacctccctg 3194
Db 523 GCTAGAGTTCTCTGAGAAATCCTCTGTAGCCATGACATTAACCTCAATCGAACCT----- 576
OY 3195 ctgccccctgtcacctcagctgctccctccctctgtgtctccacacacctccagag 3254
Db 577 TTTTATTAATGAAATGCTGATACCTGCTCCAAACATTCCTTGTCTTCCATCTTCCGAA 636
OY 3255 ctccacctctccagcacaagccccggaaccccaaggtgacagcactgtgtgaaactgtg 3314
Db 637 CTCAAGCTCCTGCTCCAGCTTCCAGACACAGAAATGCGCAGCATGTTCCATCTGCACTTC 696
OY 3315 acccttccggagcagaacacttctcactggtgctctgtgtgacggagcttggcactgacct 3374
Db 697 CGAGTACCGCGACGACGACTTCTCACCGGCTCTCTTACCAAGAAATGCGCTGCTGCT 756
OY 3375 cgaacctgaagcttgaaggcgatctcctgtgtgacagaagaagccaatgctgtgacag 3434
Db 757 GGATGCCGAAGGGAAGGAATCAACCAAGTACAAAGAAAGCTGTCAATTCACAG 816
OY 3435 cctgtctatgtgccaatgacactgacaccccgctacgcccgaagcgacacttgaagcgtcgt 3494
Db 817 CCTGCTAAGTTCTTACGACGCTGACCCACGCTGTGTCAACCAAGAGTGAAGTCAAAAT 876
OY 3495 ggcgcagcttactctgacactgcttctcgtatgtgacgggataccttgcacgcgtgtcat 3554
Db 877 CGCGGCTTCTTACTACTCTTAACTTGTGATCATTTTGTGATGCTTGTGCCACAGCTGTGTA 936
OY 3555 cttgtctgaaggcccaaggtcagcggtcaagacttgcctcaatgtctgactcagaacaga 3614
Db 937 CTTTACAGTTGCAATACTGCGAATACCGCA-----CCAGTGGCTCGGA 981
OY 3615 aggcgaagggaacatttgggggtacacacacccctctgtgtgacatgtgccaatgtgtgtg 3674
Db 982 TGAAGAACAGAAAGAGACCGGCTGCATTAAACAGAAATGTGCTGTGCCATACAGGAA 1041
OY 3675 ccccttagccccctgtgctccggcgccagcatctcccaaggggcacaacggcttctcgcg 3734
Db 1042 TAAATTTAATTTTAAACAAAGTGAATAGT-----GCTGTCTTCTTGTGCCCTATTAAGA 1095
OY 3735 aggcgtgtccctctgtgtgacgaagccggacacttgtgtgctgtgtgtgtgtgtgtgt 3794
Db 1096 GTACAAACATGCTGAAACGCGGACACTACTGCAACCTCATGATGTGCTTCTCTGATCAT 1155
OY 3795 gaataaacaacgagcgcgtcctctgacgcgtgtggccaactgtacacttccccagct 3854
Db 1156 GAAATAAGCTGATCAGAGCTCATTTAGGAAGTGTGATGTGACCTCCATCAACGACGCT 1215
OY 3855 ggaagcgtgtgtgactgtgtgacttgcctgctgctgtgtgtgtgtgtgtgtgtgtgtgt 3914

Db 1216 CAACAGCATTTAGANCTACTTTTCATCTGTGTATTGTGTTTGAAGTATTAAGGAAACAA 1275
OY 3915 ggccttgaacgcatcaacaagcctcaacattcaaaaaatctgtgataatgaaggcgcgct 3974
Db 1276 GAGTTCTGACAAAGTACATGACCAAGTCTGCGAAGTCAAGGAGTGTCAAGGCGCGGCT 1335
OY 3975 aagaagaagccattctgtgtacccatcgagctcgacagaagaatgtgtctggtcgagtgtga 4034
Db 1336 GGAAGAGGCTTGTCTCCGTGGAGAGGGGCGAGAGGGAGATGATGCTGCGCGGGCTCC 1395
OY 4035 gaagagccgt-----ttggaaatccagaagaacgtgtgcgtgcygaagaagcgtcaaca 4088
Db 1396 AGGAAACGACCGATTTCCAGGCGCTTAATTAATAATTTGAATGAAAGAAAGACAGACACA 1455
OY 4089 ctggaagcaaacctcaacgacgcgtgtgacaagaccgaaggaatgaatgtgaacacagagcct 4148
Db 1456 TTGGCGGCAAGCTAATAGAGAGCTAGATTAACAAAGCCGAGTTTGAATCAAGAAAGCTT 1515
OY 4149 gttggaagggaaccttggcaacccgaagcgaagcctagtgtgtctgtgacacactgagatcat 4208
Db 1516 GATCAAGTGGCAATCTGTGACAGAAACACATTAATCATCTGTGATATGACGAAACAT 1575
OY 4209 cgtgcagacggtgactgtcttccagaagccggagagcgcttggggggcgagtgtgaagt 4268
Db 1576 TATCCAGGCGAGCTCGGCTCTGCACTGTAAAGACACCTGCTGGGAGGTGTCTAGAGGT 1635
OY 4269 tgtgtctacagcttggcgagtgccagagtgccctcttcttctgtgacacatgacctggcac 4328
Db 1636 GCTGTGAAATTTCTGAACTGTGATCAGATGACACCTACTGACTCATCTGTTTGCAAC 1695
OY 4329 ccagagggcccttgtgtccaaagtcccgagctgtctgttctggaggaggacagagcgtgt 4388
Db 1696 ACTCCGTGCTCTCATGCGCAAGTTTGAAGCTTACTCTTTCGAAGAGGAGGTGAAACATGT 1755
OY 4389 tgcgcacccgtgtgaggctctcactcagacactgtgtgcccagcgcacacacatccgcac 4448
Db 1756 TTTCGACCTATGACCAAGTCTGTGACACCTGACGACGACACATGATGTACCCGGAG 1815
OY 4449 gcaagcagagcgtcgtgtgtactgtcatgtgcagagaacttgcagatcgaccacactt 4508
Db 1816 CCAGGCTGTGCAACCTTACTCTCTCATGTAGGTTTGGAGGACACCAAGTAATTT 1875
OY 4509 tgcgcgtgtgaaagtacagatcacatgtctctcgtccctgtgtgtggaggagcagcaga 4568
Db 1876 TGAAGAGTAAAGATGCCAAGTAAACATGTCTGCGATCTTGTGGGAAAGCACACGA 1935
OY 4569 ctctagtgaaagacacctgtgcagcttcaactcaaaacacatcctcaactatgtctgaagaga 4628
Db 1936 CTTTAAATGAAGAGCACCTGAGAAGATCTTGTAGACAATTTTGGCCTATTCAGAAAGCA 1995
OY 4629 catggtgctgcygagcagacacttgcagagaggttccaagacctgtatgtccaacttga 4688
Db 1996 CAAAGCATGACAGATGACTCTTTTCCACCCAGGTGAGGAACCTTCTGTAACTTGA 2055
OY 4689 catgatcctgaagcagaacggttgaagatgaagaaacacaggaagcctctgagatgcat 4748
Db 2056 TAGCATCTTATATGACACAGTGAATGAAGGAATTTAGGAAGATCTGTGAGATTCCTAT 2115
OY 4749 cgaactcatgtacagaatttgcgggggttaccaggggttcacccggaacttgcgtgtgacct 4808
Db 2116 GGATCTCATGTACAGAAATTTGCCAAGATTACAGGACATCTCCTGATGCGGCTGACTG 2175
OY 4809 gttgacaacaatgycgcgggaagcagcggaagcttggcaacacacgcgcggagcccgagtg 4868
Db 2176 GCTTCAGAAATGCGAGAAACACACCAAGAAAGTGTCTTACAGGAGGCTGACATGTG 2235
OY 4869 catgtgtcacgcgcgcgcctctgtgtgagtacctgcgccttctgtgacagcagcagca 4928
Db 2236 CCGTGTGACGCGCGCTGTGAGTGTGCTGAGATATCTGAGCATGCTGAGAGCACACACTA 2295
OY 4929 cctgcgcgtgtgtgtgtgttcttccacagacatcctcaacagtgctagaagagtcgcg 4988
Db 2296 CCGTGCCTGTGGCAGTGTACAGCTTCCAGAAATATTTCTTCAATGTGCTGAGAGTGTGT 2355

[illegible]

Db	1322	TACCAATCTCCCTGGCATCTTTTGGTGGAGAGCACACAGACTTTAATTAAGACACTGTA	1381
QY	4589	gaacttactcaaaacacatccctaccatgactgctgagagagaaatgggctcggagacga	4648
Db	1382	GAAGATCCTTGAGAGCAATATTTGGCTATTGAAAGAGGACACAGCATGAGATGATC	1441
QY	4649	ccttcgcaagagcaggtccagagacctgatttcaacctgcacatgatcctgcagcaag	4708
Db	1442	CTTTTCCACCCAGGTGGAGGAATCTCTGTAATCTGTAATAGCATCTTAATATGACAG	1501
QY	4709	tgaagatgaaagaaacacagagagaccctgagatgcatacgcacaaatgtaagaattg	4768
Db	1502	TGAAATATGAGGAATTTACAGAGATCCGAGATGCTTAATGATCTCATGTACAGAAATTG	1561
QY	4769	ccgggggtaccagaggtctcaacgaccccttcggctgcactggtgtgcagaacaatggcggga	4828
Db	1562	CCAAGAGTTACACAGGCATCTCTGATCTCGGCTGACTCGGCTCCAGAACTGGCAGAGA	1621
QY	4829	agacagcggagctgtagaacaacagccggagggcccgcaatgtaattgtgcagccggccccc	4888
Db	1622	AACACACCAAGAAAGAAAGTGTACACGAGAGCTGCCATGTGCTGCTGACCCGCTTGCT	1681
QY	4889	tcgtggtgtagtacctcgccctgcgtctgaaagaaacagcgcacactgcacgctgggtcgtt	4948
Db	1682	TAGTGGCTGAGTATCTAGCATGCTGGAGACACACAGCTACTCGGCTGGGCGAGTGTCA	1741
QY	4949	ccttcagagaacatctctcaaacgctgctgagagagttccgcatctccgaagaacctgt	5008
Db	1742	GCTTCCAGAAATATTTCTTCCAAATGTCTGAGAGATCTGTGGTCTGTAGAGACACCTGT	1801
QY	5009	cgcccgacggagagaggtcttcgtcccggaagaacacttaacagctggggtgtgtagagt	5068
Db	1802	CACCTGACGAGAGATGGGGGTGTGGCGAGGCCAGTACTTCCAGAGAGTGGCCTGTAGCC	1861
QY	5069	tgtctggaacaggcagccggtacttaaccatgaggcgggctctagagggcggltgaatgag	5128
Db	1862	TCTGTGGAGCGGCGCGGAGACTCTTTGACGCGGAGCGCTTATATGACAGATTAAATGAG	1921
QY	5129	ttctaaagaacatccatcccatctgtgaagccacagctgactacaagaagctggccgg	5188
Db	1922	TCTACAAAGCTGTGATCCCATCTCTAGAGGCCATCGAGATTTCCGGAGACTGACATCA	1981
QY	5189	tgcacggcaaatgcagagagccttaaccaagatcaatgcacagagttccggtcggagac	5248
Db	1982	CTCACACCAAGCTGCAGAGAGCCTTTCGACAGCATGTTAAACAG-----GATCATTAAGA	2035
QY	5249	gcggttcggagaaagtattcccggtggggtcttaagcggcccaattggtgactgagt	5308
Db	2036	GAATGTTTGGAACTTACTTCCAGTTGGTTTCTTTTGGATCCAAATTTGGGGATTTTGGATG	2095
QY	5309	agcagagcttgtgtcaaaagagacatgcatacgaagaactgycagagatctcaaccggc	5368
Db	2096	AACGAGGATTGTCTCTCAAAAGAGCCTGCATTAACCAACACTCTTCGAGATCTCCATAGAC	2155
QY	5369	tggagaggtctctcaacggagagatttggagagacgttcgttgagatatcaaaagactctt	5428
Db	2156	TAGAGGATTTTATGCTCATGTGTTGGGAGAGAAATTTTGGAAGTATTAAGACTCCA	2215
QY	5429	aacctgtgacaagtccaagcttgcatacacaagagccataccagatcacgtatg	5488
Db	2216	CTCTGTGGACAAACCAAGTTGGATTCCTAACAAGCCTTAATACATCACTTTTGTGG	2275
QY	5489	aacctgattcttgatacctaagagctcaagaagccgggtgacctactttgaccgaactatg	5548
Db	2276	AGCCCTACTTTGATGTAGTATGAGATGAAAGACAGGGTCCATCTATTGGAACAAATTTTCA	2335
QY	5549	ggctcgcacatcccggtctcgaagccgttcaagccgttcaagcggatggcgcaagaaggagc	5608
Db	2336	ACCTCCGAGAGTTCAATGTACACCAACCCGTTTACCTTGAAGGGCGGCTTCGGGGAGAGC	2395
QY	5609	tgcgcggaacaacaagcgtlaagagcgtctgaagacacgacacagccttccctatacata	5668

Db	2396	TGCTATGAGCAGTATACAGAAAGAAACACAGTCTTGCACACTATGCACCCCTTCCCTATCATCA	2455
Qy	5669	agactgcgcatccgltgtgtccaccggyagagaacgltgtcgaagccagltgagyltgagcca	5728
Db	2456	AGACACAGGATCAGCGTCATCCAGAAAGAGAGAGCTTGTGTTGACACCGATTGAAGTTGCCA	2515
Qy	5729	tcgagagcatgcggaagaagacacggaacgtgcctcttgcaccgagacagagaccacag	5788
Db	2516	TTTGAGACATGAAGAAAGAAACCCCTCACTATGACATGCCATTAAACAGAGGCGCGCTG	2575
Qy	5789	atgcctaagatgtctacagatgtgtcttcacaggtcgtctlaagagccaccgltgacacaggtc	5848
Db	2576	ATCGAAAGATGCTTACAGATGTGTCTCCAAAGGCTGTGTGGACACTACTTAATCAGGAC	2635
Qy	5849	ccctgtgaggtgtgcccaggtgttlltagcagagatccccggaagaccacgaactctccgcg	5908
Db	2636	CACGTGAGATGAGCCCAAGTGTGTTTTGGCTGAATTCCTGCTATCCAAACACTATACGC	2695
Qy	5909	atccacaacaattgagcgctctgtcttaagaactctgcagagaatgtagatgcgcgc	5968
Db	2696	ATCCACAACAAGTTGAGGTTATGCTTTAAGGAATTCATCTGAGATGTGTGAAGTGTAG	2755
Qy	5969	ggaanaaa laaagccctgtatltggtgcgagacagaagatgataccacgttagctggaagcca	6028
Db	2756	AGAAAAACAAGCGTCTCATCAGCGCAGACACAGGGAATATCAGCAGGACTCAAAAAGA	2815
Qy	6029	actactgcgcctctgcggagaggtctgtcagagccctgtctacccagcgctgtcccacgtg	6087
Db	2816	ACTTAAACAAGCTAAAAAGAAACCTCAGGCCAATGATCTCAGGGGAAAAATTCAGAAACTG	2874

Search completed: October 5, 2002, 07:35:44
Job time: 18981 sec

Db 840 CCTCCCTTACCGCTTATGCGCCAGAACTTGAGATTGGCATAACTTTGGCCGTGTA 899
 Qy 4520 agatgaggtatcaatgctctctctgcttggtgggagcagcgaacttcaatgaag 4579
 Db 900 AGATGCTGGTGAGACATGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 959
 Qy 4580 agcactgtgacgttcaactcaaaacatccatcaactatgctgaggaagacatgagctg 4639
 Db 960 AGCATTTGAGAAAGTCCCTCAAGACCTCTGACCTACAGAGAGAGGACATAGGGCTGA 1019
 Qy 4640 gggagagcactcgcagagacaggttccagactgatttcaactctgacatgattctga 4699
 Db 1020 GGGACAGCACTTGGCTGAGCAGCTCAGAGCTCATGTTCAACCTGACATGATCCTGA 1079
 Qy 4700 cggagacgttgaatgaatgaaggaacacacagagagagactatgactatcagacactatg 4759
 Db 1080 CACACACGGGAGATGAGAGAAACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1139
 Qy 4760 acagaattgcccgggagctaccagagctcagcagacttgcgtgacatgattgacagaca 4819
 Db 1140 ACAGAAATGCTGCTGAGGATACCAAGGCTCTCCAGACCTGGGCTGACATGCTGCAACAA 1199
 Qy 4820 tggcgggaaagacagcggagagctgagcaacacagcggagcggcggatgactgagtcagc 4879
 Db 1200 TGGCTGGCAAAACATGACAGACCTGGCAATCATGACAGGCTGGCCGATGCTGCTCAGG 1259
 Qy 4880 cggcggcctcgtgctgagtagtaccctgctcagagacacagcggcggcggcggcggcgg 4939
 Db 1260 CCGCTGGCCCTGCTGGCGGATACCTTGGCTCTCTGAGGAGCAGTCCGACACTGCTGTGG 1319
 Qy 4940 gctgctgttccctccagagaatctcaatcaagtgctagagagctgcgacatctcagag 4999
 Db 1320 GCTGTGTCTCTTCCAGAAATGCTCATCAACGCTGAGGAGTGTGCTGATGCTGATG 1379
 Qy 5000 acatcctgtcgc 5059
 Db 1380 ACATCTGTGCTGGCAGATGAGAGAGGGCTTGTGCTAGGAAAGACTTACAGAGACTGGGTC 1439
 Qy 5060 tggtagaggttgcctggaacagagcggcgtacttcaacatggcggtctcagagcgag 5119
 Db 1440 TGGTGGGGCTGCTGGAGACAGCGCGCGGCTACTTCAACATGGGTGCTGTATGAAGAGCG 1499
 Qy 5120 tgaatgaggtctcaaaagaactcatcccaactctggaagcccaacttgaatcaaaagc 5179
 Db 1500 TGAACGAGTCTTACAAAACCTTATCCCATCTTGAAGCCCAAGAGACTAAGAGAGAGC 1559
 Qy 5180 tggcggcgtgacagc 5239
 Db 1560 TGGCTGGCGGTGACAGGGAACCTGAGAGGCTTCAACCAAGATTATGACACAGAGCTG 1619
 Qy 5240 gctggagagcgctgctcggagacgtattccgctgaggtctcagcgccgacactcggtg 5299
 Db 1620 GCTGGGAGCTGTATTTGGGACATATTTCCGAGTGGGCTTCTATGACACAGATTTGGTG 1679
 Qy 5300 acctgagtgagcagaggttgggtgtaaacagagcagatcagagagcctgagagagatct 5359
 Db 1680 ACCTGATGTAACAGAGATTTGTGTACAGAGAACCGTCAATCAAGAGACTTGGCAGAGTCT 1739
 Qy 5360 cacacggcgtgagaggttctacacgagagatttggcgcgcgcgcgcgcgcgcgcgcgcgc 5419
 Db 1740 CACACCGCGGTGAGAGAGTCTTATACGGAAGTTCGGGGATGATGTGTGAGATCATCA 1799
 Qy 5420 aagaacttaccctgtgagacaaagtccaagcttgaactcaaaagagcctacatccagatca 5479
 Db 1800 AAGATTTTAAACCCAGTGGACAAAGTCCAAAGTCCCAAGAGAGGCTATCATCAAGATTA 1859
 Qy 5480 cgtatgagagcagcttctgatacctcaagagctcaagagcgggtgagacacttggacc 5539
 Db 1860 CCTATGTGGAGCCACATTTTGACACTTATGAGCTCAAGAGTCCGGGTGACTTGTGATC 1919
 Qy 5540 gcaactatgggtctgcacattcctgtctgcagcgcttcaagcgagatggcgcgac 5599

Db 1920 GGAACATAGGGCTGCGGGCTTCTCTTCTGACACACCTTACACACGATGAGCTGCC 1979
 Qy 5600 acgggagctgcccaggaacacacagcgtlaagacgctgctgaagcaccagccttcc 5659
 Db 1980 ACGAGAGATTGGCCGAAACAGCAACAGCAGCGTGTGAGACAGCAGAGACTCTTTC 2039
 Qy 5660 cctacatcaagactcgcgtctgtgtgccacgggagagagacggtctcagacccatgg 5719
 Db 2040 CCTACATCAAGACAGCAGTCCGAGTGTCCACCGTGAAGAGACAGTGTCTGACACAGTGG 2099
 Qy 5720 aggtggccatggagacatgacagaagaagacagggagctggtcttggccacggagcag 5779
 Db 2100 AGGTGGCCATTGAGGACATGCAAGAAAGAGACCGGAGACTGGCTTTGGCCACGAGCAGG 2159
 Qy 5780 acccaccagatgctaagatgctaagatgcttcaagggtcgtlaaggccacgtga 5839
 Db 2160 ACCCTCAGATGCCAAAGATGCTGCGATGATGCTTCCAGAGCTTCTGTGGGAGCCACTGTGA 2219
 Qy 5840 accaggtcccttgagagtgccagagtggttttagagagagatccgggaagaccgaagc 5899
 Db 2220 ACCAGGTCCTTGGAAAGTGGCCAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2279
 Qy 5900 tcttcggacatcac 5959
 Db 2280 TCTTCGACATCAACAAACAACTCCGGCTCTGCTTCAAGATTTCTGCAAAAAGTGCAGG 2339
 Qy 5960 atgctgtcgggaaataagagccttgatggccggagacaaagagataccacgctgagc 6019
 Db 2340 AAGCAGCTGAGAAAGAAAG 2399
 Qy 6020 tggagcgaactgacgtccgctcgggagagctctgacagcgcctcttaccagcgtgc 6079
 Db 2400 TGGAGCGTACATATAGCCGCTCGCGGAGGCTGTGACACCTCTGCTTACCAAGTGTGC 2459
 Qy 6080 cccagctatggtgacccacccacccacccacccacccacccacccacccacccacccaccc 6139
 Db 2460 CCGAGCTGCTGGCACCAAGTTCCACACAGCTCAGAGAGCTCCATGACAGATCAAGTTTCA 2519
 Qy 6140 gaaagcagagcctgtgagcccaagagacaaagctgacagtagagagacagacggcg 6199
 Db 2520 GGAAGGCTGACCTGTG-----ACAAGCTAAGAGGCACACACAGACAGACAGACCCAC 2574
 Qy 6200 gctcagctgtctgtgc 6216
 Db 2575 ACCTCAGCTGCTGTGC 2591

RESULT 2
 BC020473/C 1510 bp mRNA linear HTC 03-JAN-2002
 LOCUS BC020473
 DEFINITION Homo sapiens, clone IMAGE:3853956, mRNA.
 ACCESSION BC020473 GI:18042973
 VERSION BC020473.1
 KEYWORDS HTC.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1510)
 AUTHORS Strausberg, R.
 TITLE Direct Submission
 JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
 USA
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 COMMENT Contact: MGC help desk
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: ATCC
 cdna Library Preparation: Life Technologies, Inc.
 cdna Library Arrayed by: The I.M.A.G.E. Consortium (ILMIN)
 DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center

AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	High efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2 (sites)
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
PUBMED	Genome Res. 10 (10), 1617-1630 (2000)
REFERENCE	20499374
AUTHORS	11042159
TITLE	3 (sites)
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, Y., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasliwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawal, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
PUBMED	Riken integrated sequence analysis (RISA) system--384-format
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	11076861
TITLE	4 (sites)
JOURNAL	The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
PUBMED	Functional annotation of a full-length mouse cDNA collection
REFERENCE	Nature 409, 685-690 (2001)
AUTHORS	5 (bases 1 to 1507)
TITLE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Aikawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, Y., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Tejima, Y., Toyota, T., Yamamura, T., Yamanka, I., Yasunishii, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
PUBMED	Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-72 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
REFERENCE	10349636
AUTHORS	10349636
TITLE	2 (sites)
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
PUBMED	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	11042159
TITLE	3 (sites)
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, Y., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasliwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawal, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
PUBMED	Riken integrated sequence analysis (RISA) system--384-format
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	11076861
TITLE	4 (sites)
JOURNAL	The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
PUBMED	Functional annotation of a full-length mouse cDNA collection
REFERENCE	Nature 409, 685-690 (2001)
AUTHORS	5 (bases 1 to 1507)
TITLE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Aikawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, Y., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Tejima, Y., Toyota, T., Yamamura, T., Yamanka, I., Yasunishii, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
PUBMED	Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-72 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
REFERENCE	10349636
AUTHORS	10349636
TITLE	2 (sites)
JOURNAL	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
PUBMED	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	11042159
TITLE	3 (sites)
JOURNAL	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, Y., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kasliwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawal, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
PUBMED	Riken integrated sequence analysis (RISA) system--384-format
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	11076861
TITLE	4 (sites)
JOURNAL	The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
PUBMED	Functional annotation of a full-length mouse cDNA collection
REFERENCE	Nature 409, 685-690 (2001)
AUTHORS	5 (bases 1 to 1507)
TITLE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Aikawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F., Hume, D., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, Y., Kojima, Y., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Schiraldi, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, Y., Tejima, Y., Toyota, T., Yamamura, T., Yamanka, I., Yasunishii, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
JOURNAL	Direct Submission
PUBMED	Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-72 Suenho-cho, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
REFERENCE	10349636
AUTHORS	10349636
TITLE	2

Query Match	Best Local Similarity	Score	DB	Length
Matches 1092; Conservative	78.5%;	Pred. No. 6.3e-146;	Mismatches 206; Indels 93; Gaps 1;	
QY 1670	tgtagccgacagcagctcaacttcaagcagcgcagaggtccgtgacgaacctgtgtgac	13.6%;	DB 11;	Length 1507;
Db 1	TCGACCCACACAGCCCTCAATTTGACAGAGCGCCAGGCGCTGTGCGCAACGTGCGTGC			
QY 1730	gagtgacgtatcatgacagagcagagcccaagcagagcttgcgcgtatcctttgcaagt			
Db 61	GGATCCAGTACATGCGCGGTGAAGACCAAGCGCAGCCCTTGGCGGTCATCTTTGGGAAT			
QY 1790	ccagctgcagtaatttaccgcgcgcgcgccttaaccgcggtgtgttaccatacaagctcc			
Db 121	CTAGCTGACGCCAATTCACCCGAGAGGCGCTTACACCACTGCTTATCAACAACATCTTC			
QY 1850	ccgagttctcaagagagttcaagctgcacatcttccagcctctgcgtgacagagaacatacc			
Db 181	CGGAATTTACAGAGGAATTCAGAGCTACGACTTCCGCGCGTGAGACCGAACAATCACCC			
QY 1910	tgctgttcaaccttaccatgtcaagctgcagcccgcgccgagacgtccctgtgagaaac			
Db 241	TCCTTTTTCCTTACCACTATGTACAGTGTGACGACCCCGCGCAGAGAACAGCCTTGAGACGC			
QY 1970	ccgttgagcttacttggatcccaatgcgttcagcaagggcgcccttgagagaccgcccctt			
Db 301	CTGTGGGCTTCACCTGGATCCCTCTGTTCACAAATGCGCCGCTGAGAGATGTGCTCTCT			
QY 2030	gtctcccaagtgtctgtgagacagccgcgcgcgcagctatctcgtgtctcaacccgaatgtg			
Db 361	GGCTGCGCCGTGCGGTGAGACCAAGCCTCCACCCAGACTGCGTCTGACGCCAGATGTAG			
QY 2090	cgcttccgagcatgcgctgtgtgacagctacaaagggcglttcaagtgtgagagctcaag			
Db 421	CGCTGCTGTGCATGCGCTGGTGTGAGCGCCACAAAGGCGSTGTTCACTGATGTGAGCTCACCG			
QY 2150	ccgtgtctctgtgtgaaccccccaagagacccttacctgagacaattcttcaacctgtgtgacag			
Db 481	CGGTGTCGTCCCTGACACCGCGAGAGACCCCACTTGGATTAAGTTCCTTCAAGCTGTGATCAAG			
QY 2210	tccctgagagagagagccttcccatcccgctcaagaaacgtgtgtgagcgaagggcaag			
Db 541	TTCTGAGAGAGAGGATCTTCCGCTTCCGACTCAAGAGAGCACTGCTGAGACTTAAGGCACTA			
QY 2270	tggaagcagagagctgtgagggcagcttcttgcagcaactgcgctgtgcccagccccgaacctgtg			
Db 601	TGGAGAGAGAGTGGGAGCGCAGCTGCGAGCCCTGGCGCTTCCGACGCCAGAGGCCCTTAG			
QY 2330	tgagcttctccacacgtgtgtgacaaagctcgtgtgtgtgtatcaagcccccgatca			

Db 661 TAGCCTTTCCACCTGCTGTAGACAACTTTCGCTTGGTTGTCGGGCCACCCATCA 720
Qy 2390 tcaatgagccagatgtgaacctgagccgtgagaccttgaagcaatggccatgta 2449
Db 721 TTTGTGGCCATATGTGTAACCTGGGTGAGAGGSCCTTTGAACCTATGTGTAAGCA 780
Qy 2450 gecttgcacagagccctgagagcaagccagagatgcccggtacatgtccacagctg 2509
Db 781 GCGTTGTGCACCGGAACCTGAGAGCTGCCAAGACTCCCGGGCCACTGCGCACTGCTG 840
Qy 2510 ctgactgcgtccactagacaccttgcctctctgtgcaatgtgcccagctccgagatgg 2569
Db 841 CTTCTATGTCTACTAGCCCTTTCGCTCCCTGGGTGTGACTGTAGCCTCCAGGCGAAG 900
Qy 2570 cccctcagtgacagtgacagctgacacactgagccgtgagctgtgtgcgcccagagcc 2629
Db 901 CCCCTCCAGCAACTGTGAGAGCTGCCAAGCTGGCCCGTGGCTGTGTCGCCGCCAGCC 960
Qy 2630 tctactgtgagcggttccaaagagcatcagcagcaagcaacctgacctgcgtgagccctg 2689
Db 961 TGTACCTGGCACGGCTTACAGCATCAGACAGCAACCTGACTGTGTGCTCCCTG 1020
Qy 2690 gctctgtgagatgagagaggttcccgacatcctctgacagc----- 2727
Db 1021 GCTGTGTGAGAGATGAGGTGTCCCGCATCTTACGCAAGCAAGCCATGACCGCACTTA 1080
Qy 2728 ----- 2727
Db 1081 GCGCGGCTCCTCTTACCTGAGAGGCTCATCTCGGCTCGCCACGCCCTCACTCGAGAC 1140
Qy 2728 -----aagctgtctcaagagagagctgtgctgtgagtggtgtgaagcaagt 2776
Db 1141 ACACTGTGCAAGAACTGCTCTCAGAGAGAGTGGCTGTGCAAGTGGGTGTAGCGGAGT 1200
Qy 2777 ccgtaacgagagccatccctccacagacgcgtgtctctctccagctcaatgtgtgaagat 2836
Db 1201 CGGTGTGAGAGCTAGTCTCAGACAGATGCTGTCTTCTTCCAGCTCATGTAAAAAGCA 1260
Qy 2837 tgcgctgtcacctgtgctgtgagcagcagctagacacaccccgaaagtgtgctcccg 2896
Db 1261 TGGAGCTTCATCTGCTTCTGGGCGAGGAGACTGCTCCCGCAACGATGCTCTCCCTG 1320
Qy 2897 gacgctccctgagagacataactgacctgtgtgagctctgtgagcctgagaggtatatacc 2956
Db 1321 GGGCTTCTCTGATGATGATTCGCGCTGTGTGGCTTCCGTTGAGGCTGGAATCATACCA 1380
Qy 2957 gtgtccacaag 2967
Db 1381 GAGTCCATAG 1391

RESULT 4
BM473590 1112 bp mRNA linear EST 05-FEB-2002
LOCUS BM473590 6466302 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5562002
DEFINITION 5', mRNA sequence.
ACCESSION BM473590
VERSION BM473590.1 GI:18522632
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1112)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be
found at the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLNL2290 row: 1 column: 01
High quality sequence stop: 559.
Location/Qualifiers
1. 1112
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5562002"
/clone_id="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: PCMV-SPORE6;
Site 1: NotI; Site 2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH-MGC Library."
BASE COUNT 223 a 295 c 325 g 266 t 3 others
ORIGIN
Query Match 12.5%; Score 797; DB 10; Length 1112;
Best Local Similarity 96.8%; Pred. No. 1.3e-133;
Matches 867; Conservative 0; Mismatches 21; Indels 8; Gaps 5;
Qy 5484 tgtgaaacggtacttgaactc---acgagctcaagagagcgtgacctattga-c 5538
Db 916 TGGGAACCGCTACTTGTGATTAATCTAGCAAGCTCCAAAGAACCCGGGTGACTTATTGACC 857
Qy 5539 cgcacatatygc-ttcgcacatcctgttctgcacgaccttcaag-ccgagatggcgcg 5596
Db 856 CCAACTATGGGCTTTCGACATCTCTGTTGGACGCGTTCAGCAGCCGATGGGCGCG 797
Qy 5597 ccacagggagagctgcccagagcaacagcgttaagagctgtctcaagcagcaagcgcct 5656
Db 796 CACAGGGAGGTGCGCCAGCAACACAGCGTAACATCCTCTCAGCANACACAGCGCT 737
Qy 5657 tcccc-tacataaagactcgacatccgtgtgtgtccacggagagagcgtgtctagcgca 5715
Db 736 TCCCCNTACATCAAGACATCGCATCCGTGTGTGCCACGGGAGAGAGCGTCTACGGCA 677
Qy 5716 gtgagagtgagcactcagaggaatcagagaagaagacagggagctgtgtccacgag 5775
Db 676 GTGAGGTGGCCATGAGAGACTGCAGAAAGAACCGGGAGCTGTGCTTGCACCGAG 617
Qy 5776 caagaccacacagatgctaaagatgtacagatgtgtgtctcaagggctctgtatgggcccac 5835
Db 616 CAGGACCCACCAAGATGCTAAGATGCTTACAGATGGTCTTCAAGGGCTGTGTAGGCCAC 557
Qy 5836 gtgaaccagagttcccttgagaggtgagccaggtgttlttaagacagatcccggaagacc 5895
Db 556 GTGAACCAAGGTTCCCTGAGGTGGCCAGGTGTTTATGACAGAGATCCCGAAGACCCC 497
Qy 5896 aagctcttcgggcatcaacaacaatgtgcgtctgtctcaagagcttctgaagaatgt 5955
Db 496 AAGCTCTTCCGGCATCACAAACAAATGCGCTCTTCAAGGACTCTGCAACAAATGT 437
Qy 5956 gaggatgcgttcgagaaataaaggccctgtatgtggccggagcagaagaggtatcacacgt 6015
Db 436 GAGATGCGCTGCGGAAAAATTAAGCCCTGATTGTGGCGCGACAGAAAGGATACACCGT 377
Qy 6016 gagctgagcgcaactactgctgcgctcgtgagagagctctgcagccctgtctataccagcg 6075
Db 376 GAGCTGAGGCGCACTACTGCGCGCTGGGGAGGCTGTGCAGGCCCTTATACCAACGCG 317
Qy 6076 ctgcccagctgtatgagcaccacccaccccggtctcaagaaatccttgaacagagcaagt 6135
Db 316 CTGCCCCAGCTGATGGCACCCACCCACCCGCTTCAGGAATCTCTTGAACAGAGCAAGT 257
Qy 6136 ttccgaagagagagcctctgagcccaagaagcaaaagcgttactctgaggaaccagcac 6195
Db 256 TTCCAAAGGAGACCTGTGAGCCACAAAGAGCAAAAGCTGTACTTGTAGAGAACCAACGAC 197

Db 61 CTCTGCAATCTGAACAGCACTTCTGTACGACACAGTAGATGAGGGAATTCAGGAAGAC 120
 Oy 4735 cctgagatgtctatcgaccatgcatgaatgtgcccgggctcaacaggctcacggagc 4794
 Db 121 CCTGAGATGCTTATGAGCCCTATGATACGATTAATGCAAGATGATCCAGAGCATGCTGAC 180
 Oy 4795 cctcgatgtgacccgtgtgtgcgaacatgagccgggaacagcgaggtgtggcaacacgac 4854
 Db 181 CTTGGGCTGACTTGGCTCCAGAAATGCGAGAAACACCTAAGAAAGAGTCTTACAA 240
 Oy 4855 gaagcgccagtcagtgtgcatgtgacggcgccctcgtgtgagtaactcgccctgtc 4914
 Db 241 GAGGCCGCCATGTCCTGGGTCATGACAGCGCCCTGCTGTCGACGATCCTGAGCATGCTG 300
 Oy 4915 gaagacagcgccacccgtggcctgtgtgttctcttcacagaaatccatccacgtg 4974
 Db 301 GAGGACCAACAGCTACTGCTGGGTCGAGCGTACGCTTTCAGAAATTTCTTCCAAATG 360
 Oy 4975 ctgagagagtcgcacatctccgacacatctgtcgcgcgacgagagggctctgtcc 5034
 Db 361 CTTGAGAGCTGTGACATCTGTGATGACACTTGTCACTGATGAGGACGCGGTATGCTCT 420
 Oy 5035 gggaagacacttcaactgtgagctgggctgtgtaggtgtgtctggaacagcgagctctc 5094
 Db 421 GGTGCTACTTCACTGAGAGTGGCTGTGGGCTCTCTGAGACGAGCTGGAGCTCTTC 480
 Oy 5095 acatgtggcggtcctcagcgagcggtgaatgaggtctacaagaactccatccatccg 5154
 Db 481 AGCAGCGGAGGCTTGTACGAGAGGTTAATGAAGTCTACAACTGGTCACTCCATCTC 540
 Oy 5155 gaagcccaacgctgactacaagaagctggcgcggtgcaacggcaactgcagaagcgctc 5214
 Db 541 GAGGACACACAGAAATTTCCGGAAGCTGACCTCCACTCACACAACTGCAGAGAGCCTTC 600
 Oy 5215 acaagatctgacacagatctccgctgtggagcggtgtgtcggagcgatctccgctg 5274
 Db 601 GATTAACATCAACAAG-----GACCATTAAGAGATGTTGGACTCACTTCCAGTT 654
 Oy 5275 ggtctcaagcgccacactcgtgtgacctgaggaagagagtttctgtaacaggagca 5334
 Db 655 GGTTCACGAGATCCGATTTGGGAGATTGGATGACAGAGAGTTGCTGTATCAAGAGAAC 714
 Oy 5335 tgcatacgaagctgacagagatctcaacacggctggagaggtctacaacgagagatt 5394
 Db 715 GCATACGAGATCCCGGAGATCTCATAGATAGAGGATTTATGGCCAGTGTTC 774
 Oy 5395 ggagacagctgtgtgagatctacaaagacttaccctgtgacaaagtcacaaagctgac 5454
 Db 775 GGTGCAAGTTTGTGGAAGTGAATGAAGACTTCACTCCAGTGAACAAACCAAGTTGAT 834
 Oy 5455 tcaaaaaggctacatccagatccagatccagatgtgaacggtacttatacctcagagctc 5514
 Db 835 CCAACAGAGGCTTACATTCAGATCAGCTTTGTGGAGCCTTACTTGTATGAATGTGAGATG 894
 Oy 5515 aaggaacgggtgaactacttgaacgaactatgggctgtccacatctcgttctgcag 5574
 Db 895 AAAGACCGGGGTGACTTCTGAGAAATTTCAACTCCGAGGTTTCATGTACACCACC 954
 Oy 5575 ccgttcaacggatgtggcgcgacagcggtgagctcccgagaacaaacgaagctgaagc 5634
 Db 955 CCCTTACCTCTGAGGGGAGACCCCGGGGAGACTTCAATGAGCAACACCGCAAAACACC 1014
 Oy 5635 ctgtcagacagcagcagcgtccctccatcaacagactcagctgtgtgtccacgg 5694
 Db 1015 GTGCTCACCAACGACAGCGCTTCCCTACATTAAGACAGAGTCCGAGTCAAGCCAGAAA 1074
 Oy 5695 gaaggaacggtgtgacgacgagtggaagtggtccatcgaggaacatgcagaagaacag 5754
 Db 1075 GAGGAGTTCGTTTGAATCCGATTTGAGGTTGCCATTTGAAGATGAAGAAAGAACCTG 1134
 Oy 5755 gagctgagcttgcacagcagacccaacagatgtcaagatgtcaagatgtgtctt 5814

Db 1135 CAGTTAGCCGTGGCCACTCACAGAGGCCCTCGATGCAAAAGATCTGCAAAATGACTG 1194
 Oy 5815 cagggctctgtagggccaccgctgaacagggtccctcgagagtggtgcccaagtgcttta 5874
 Db 1195 CAGGCTCTGTAGAGCCACTGTAAATCAGGACACTGGAGGTGGCCCAAGTGTCTTG 1254
 Oy 5875 gcaagatcccggaagaccccaagctctccggatcacaacaatgtgcgtctc 5934
 Db 1255 GCTGAATTCAGCTGACCCCAAGCTCTACGACATCAACACAGCTGAGGTGTGCTTC 1314
 Oy 5935 aaggaactctgaagaatgtgagatgctgtcggtggaataaataagccctgattggcg 5994
 Db 1315 AAGGATTCATAAATCGATGGGAGAGGCCGTGAGAAACAGCCGACTATCACCGCA 1374
 Oy 5995 gaccagagagatcacaccgctgagctgagcgcaactactcgcgcgtggagagctctg 6054
 Db 1375 GAGCAGCGGAGTACACAGAGAGTGAAGAAAGACTCAACACAGCTGAGAGACAGCTC 1434
 Oy 6055 cagccctgtctaccagcgctgcccagct 6086
 Db 1435 AGGCCATGATGAGCGGAAATCCAGAGCT 1466

RESULT 6
 BI770146 796 bp mRNA linear EST 25-SEP-2001
 LOCUS 60305332P1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5203063 5',
 DEFINITION mRNA sequence.
 BI770146
 ACCESSION BI770146
 VERSION BI770146.1 GI:15761724
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: rgs@bgl-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LAM11508 row: p column: 08
 High quality sequence stop: 793.

FEATURES

source
 1..796
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5203063"
 /clone_id="NIH_MGC_122"
 /lab_host="DH10B"
 /note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
 site.1: NotI; site.2: EcoRV (destroyed); RNA source
 anonymous pool of 24 week female lung, 16 week female
 spleen, and 20-22 week male spleens. Library is oligo-dT
 primed and directionally cloned (EcoRV site is destroyed
 upon cloning). Average insert size 1.4 kb, insert size
 range 1-3 kb. Library is normalized and enriched for
 full-length clones and was constructed by C. Gruber
 (Invitrogen). Research Genetics tracking code 026. Note:
 this is a NIH-MGC Library."

BASE COUNT 188 a 228 c 231 g 148 t 1 others
 ORIGIN
 Query Match 12.0%; Score 762.4; DB 10; Length 796;
 Best Local Similarity 99.6%; Pred. No. 2.2e-127;

Matches	785; Conservative	0; Mismatches	1; Indels	2; Gaps	2;
QY 5275	gagcttctacgagccaccactcgttgacccgtgagtcgacgagagttgtgtacagagacca	5334			
Db 1	GGCTTACGGGCGCCACTTCGGTACCTGGATGAGACAGAGATTGTGTACAAAGGACCA	60			
QY 5335	tcgatacgaagctgtgcagagatctacacccgctgagagagttctacacgagagattt	5394			
Db 61	TCGATACAGAAAGCTGGCAAGATCTCCACCGGCTGAGAGATTCTTACACGAGAGATT	120			
QY 5395	gagcagcagctctgtagattatcaaaagctcttaccctgttgacaaagtcacagcttgc	5454			
Db 121	GGGAGAGAGCTGCTTATGATATCAAAAGACTTAACCTGTGACAGACAGTCCAACTTGC	180			
QY 5455	tcacaaagagcctacatccagatcagatgtgtgaaacgctacttgcatacctagctc	5514			
Db 181	TCACAAAGAGCCCTACATCCAGATCAGCTATGTGGAACCTGACTTGTATCTACAGACTC	240			
QY 5515	aagcagcaggtgtacacttctgacccgcaactatggtgcttcgacatctcctgttctgcag	5574			
Db 241	AAGGACCGGGTACCTTACCTTTCACCGCACTATGGGCTTCCGACATCTCTTCTGCACG	300			
QY 5575	ccgttacgcgcggtatggtgcgcacacgagagctgcccagcaacacacgagctaaagcg	5634			
Db 301	CCGTTACAGCGCGATGGGGCGCACACGGGAGCTCCCGAGAACACAGCGTAAAGACG	360			
QY 5635	ctgtctacagacacgacagcgtctccctacatcaagaactcgcactcgtgtgtgcacgg	5694			
Db 361	CTCTCAGACACCCACACCGCTTCCCTACATCAACATCGCATCCGTGTGCCACCGG	420			
QY 5695	gagcagcaggtgtgtgacgcagctgagctgtgacatcgagagactgtgagagagacag	5754			
Db 421	GAGGAGACGGTGCTGACGCCAGTGGAGTGCGCATGAGAGACATGCAAGAAAGACACGG	480			
QY 5755	gagctgagccttgcacacgagacagacccacagatgtctaaatgtctaaaga tgtgtcct	5814			
Db 481	GAGCTGGCTTTTSCACACGAGACACACAGATGATTAAGTGTACAGATGGTGCTT	540			
QY 5815	caaggtctgtgtgagccacacgtgaaacagaggtccctctgagagtggtgccaagtgtt	5874			
Db 541	CAGGGCTGTGTAGGGCCACCGTGAACCAAGGGTCCCTTGAGGTGGCCACAGGTGTTT	600			
QY 5875	gagcagatcccgaaagaccccaagctcttcgcgcatcacaaaatgtggcctgtcttc	5934			
Db 601	GCAAGATCCCGAAGACCCCAAGCTTTCGCGCATCAACAAATTTGGGCTCTGCTTC	660			
QY 5935	aaggaattctgcagaagaatgtgagatgtgcctgtgcgaaa -aataagccctgtatgtg	5993			
Db 661	AAGGACTTTCGAAGAATGTGAGATGGCTGCGAAACAATTAAGGCCCTGTTGGGCC	720			
QY 5994	gagcagcagagagatcacacccgtgagctgtgagcgcaactac -tgcgcctgtgcgagag	6052			
Db 721	GGACCAAGAAAGATACACCGTGTGAGCTGAGCGCAACTACTTGCCTGCGGAGAGCTC	780			
QY 6053	tgcagccc 6060				
Db 781	TGCAGGCC 788				

RESULT 7
 BI831979/c 815 bp mRNA linear EST 04-OCT-2001
 LOCUS 603076025F1 NIH_MGC_119 Homo sapiens cdna clone IMAGE:5167759 5'
 DEFINITION mRNA sequence.
 ACCESSION BI831979
 VERSION BI831979.1 GI:15943529
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 815)
 AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINTL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINTL at: http://image.llnl.gov/plate/llnl1417_row_a_column_08 High quality sequence start: 25 High quality sequence stop: 812. Location/Qualifiers 1. 815 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5167759" /clone_id="NIH_MGC_119" /tissue_type="medulla" /lab_host="DH10B" /note="Organ: brain; Vector: pCMV-SPORT6; Site:1: NotI; Site:2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH-MGC Library." BASE COUNT 140 a 239 c 257 g 177 t 2 others ORIGIN
FEATURES	source
QY 5533	ttgacgcgaactatgagcttcgcacatctcgttctgcagccggtcgaogccgagatggg
Db 803	TTTGACCCCAACTATGAGGCTTCGACATCTCTGTCTGCACGCCGCTACCGCGATGGG
QY 5593	cgcgacacgagggagctcccgagcaacacagcgtcagcgtcagacccagac
Db 743	CGCGACACGGGAGCTCCCGAGCAACAGCGTAAAGCGTCTAGCACCGACCGAC
QY 5653	gacctccctaca -tcaagactcgatccgtgtgtgcacccggagagacagctgtgac
Db 683	GGCTTCCCTTACAGTCAAGACTCGATCCGTGTGTGCCACCGGAGAGACGGTGTGAC
QY 5712	gccagtgaagtggtgcatcgagacatgcagaagaagaacacgagagctgtgcttgcac
Db 623	GCCAGTGAAGTGGCCATCGAGACATGCAAGAAAGACACGGGAGCGTGGCTTGGCAC
QY 5772	cgagcagacccacacagatgtctaaatgtctacagatgtgttcaagggtctgtgagggc
Db 563	CGAGCAGACCCACCAAGATGTAAAGTCTAACAGATGGTGTCTTAAGGGCTGTAGGGCC
QY 5832	caccgtgaacacaggtgccctggaggtgtgccaggtgttlltaacagagatcccggaaga
Db 503	CACGCTGAACAGAGGTCCCTCGAGAGTGGCCAGGTGTTTATACAGAGATCCCGGAGA
QY 5892	ccccaaagctcttcgagcatcaacaagaatgtgcgtctgtcttcaagagactctgcaaga
Db 443	CCCCAAAGCTTCCCGGCATCAACAACAATTTGGGCTGTCTTCAAGCACTTCTCAAGAA
QY 5952	atgtgagatgtgctgtgcgaaataaagggccctgtattgtggccggaacgaagagatcca
Db 383	ATGTAGAGATGGCTGCGGAAATAAAGCCCTGATTTGGGCGGACCAAGAGAGGTACCA
QY 6012	ccgtgaagctgagcgaactactcgcgcctcgaggagagctctgacgcccctgtaccaca

QY	323	CGGTGAGCTGGAGCCGCACTACTGCGCGCTGGGGAGAGGCTCTGCGAGCCCTCTTACCCA	264
QY	6072	gcgcctgtcccccagctgatatgtagccaccaccaccaccgagctcctaagaaactccttgaacagagc	6131
Db	263	GGCGCTGCCCCAGCTGATGTGGACACCAACCCACCCGGCCTCAGGAACCTCTTGAACAGAGC	204
QY	6132	aagtttcgaaaggcagacctcttgagcccaagaagaccaaagctgtactagaggaacca	6191
Db	203	AAGTTTCCAAAGGACAGACTCTGAGCCCAAGAGACCAAACTGTACTAGAGGACCA	144
QY	6192	gacccggagccccaagctgtctgtctgcgaaaggagagtgctgccttgatgtcccaatggctg	6251
Db	143	GCACCCGGGCGCTCACACTGTCTGTCTGCTGCGAGGGAGTCTGCACTGTGTGCCACTGGGCTG	84
QY	6252	tgagggtgacacacactgtacttggggctgtggccctctgcctctgtgtcccatc	6304
Db	83	TGGGGTGACACACACTGTACTTGGGCGCTGGGCCCTGTGCCCTGTGTCCCATC	31
RESULT	8		
BI820302		927 bp	linear
LOCUS			EST 04-OCT-2001
DEFINITION		603036664T1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5178087 5'	
ACCESSION		BI820302	
VERSION		BI820302.1	GI:15931852
KEYWORDS		EST.	
SOURCE		human.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE		NIH-MGC http://mgc.nci.nih.gov/.	
JOURNAL		National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT		Unpublished (1999)	
		Contact: Robert Strausberg, Ph.D.	
		Email: cgabbs-remail.nih.gov	
		Tissue Procurement: Life Technologies, Inc.	
		cDNA Library Preparation: Life Technologies, Inc.	
		cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	
		DNA Sequencing by: Incyte Genomics, Inc.	
		Clone distribution: MGC clone distribution information can be	
		found through the I.M.A.G.E. Consortium/LNL at:	
		http://image.lnl.gov	
		Plate: LAM11443 row: 0 column: 16	
		High quality sequence stop: 806.	
FEATURES		Location/Qualifiers	
SOURCE		1..927	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/clone="IMAGE:5178087"	
		/clone_1db="NIH_MGC_115"	
		/lab_host="DH10B"	
		/note="Organ: pooled brain, lung, testis; Vector: pCMV-Sport6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb. Insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH-MGC Library."	
BASE COUNT		214 a	257 c
ORIGIN		292 g	164 t
Query Match		11.8%;	Score 751.4; DB 10; Length 927;
Best Local Similarity		97.9%;	Pred. No. 2,2e-125;
Matches 825; Conservative		0; Mismatches 11; Indels 7; Gaps 6;	
QY	3906	gggggaaaaaggccttggagcgcacccaacagcctcacaattcaaaatctctgatatgaa	3965

Db	1	GGGGAAAAAGCCCTTTGAACGATCAACAGCCTCACATTTCAAAAAATTTCTGGATVTGA	60
QY	3966	gagcgcgccagagaagaagccatctctgggtacatcgcagagctcgcgaagaatggttcgagc	4025
Db	61	GGGGGGGCTAGAGGAAGGCCATTCTGGGTACATCGGAGGCTCGACAAGAATGGTTGGGG	120
QY	4026	aagtcgtgagagagcccglttgggaatccagaagacgttcgcctggcgaagagcgtcac	4085
Db	121	AAGTCGTGAGAGGAGGCCCG- TTGGCAATCCGGAGATGTGCGCTGCGGAGAGGCTCAC	179
QY	4086	aaactggaagcaaaccttcgaccgcgttgacacaagaccgaagatgaaatgaaacaagagc	4145
Db	180	ACACTGGAAGCAAACTCTGACCGCTGTGACAAAGACCAAGATGAATGGAACACAGAGC	239
QY	4146	cttggtggaaggaagaaccttgcaaccgaggaagccagtcagtcgtctctggacacactggagat	4205
Db	240	CTTGGTGGAGAGGAACTTGCAACCGAGGCAAGCTTAATGTTCTTGACACACTGGAGAT	299
QY	4206	catctgacagacggtgtagctttcagaagcccgagagagcgtcttgggggacagtctga	4265
Db	300	CATCGTGACAGAGGTGATGCTTTAGAAAGCCCGGAGAGACGTTGGGGGCAAGTGTGAA	359
QY	4266	gatttgctgtcacaagctcgtggcagtgccccagagtgccctctctcttgacatgacctgac	4325
Db	360	GGTTGTGCTGTACAGCTCTGGGCAAGTCCAGAGTGGCCCTTCTTGACATAGGCTTGCC	419
QY	4326	caaccgagagccctctgtctcaagttcccgagagctgcctgttcgagaggaagagagct	4385
Db	420	CACCCAGAGGCGCTTGTTGTCCAAAGTCCCGAGCTGCTGTTGAGAGAGACACGAGCT	479
QY	4386	gtgtgcggaactgtgctctgaagctccctacagacactgtgacgcgcatcacaccatcg	4445
Db	480	GTGTGCGACACTGTGCTCTGAGGCTCTCTACAGACACTGTGGCAGCCGATACGACATCCG	539
QY	4446	cacgcagccagcgcgctcgtctgaactctcatctgcagagaagaaacttcgagatcgccaca	4505
Db	540	CACGCAAGCGACACGCTCTGCTACTGCTCAATGCGACAGAACTTGAGATGGGCGACAA	599
QY	4506	ctttgcgcgtgtggaagatgacaggtcaacc- atctctctcgcctcgtgt- gggggagcagc	4563
Db	600	CTTTGCCCTGTGAAAGATCAGGTCAACCAATGTTCTCTCTGTCCTGTTGGGGGAGACAGC	659
QY	4564	cagaactcagtgaaagagacacttgcaagcttcaact- aaaaacatctcctacatagtctga	4622
Db	660	CAGAACTTCAGTGAAGAGACCTCGAGAGCTTCATCAAAAACATTCCTACCTATGCTGA	719
QY	4623	ggaagaa- catgggctctgcgggaagacactctcgacagagcaggttccagaacctgatttca	4681
Db	720	GGAGAGCAATGGGGTGGCGGAGACACCTTGTGAGAGAGGCTCCAGGACCTGATGTTCA	779
QY	4682	acctcacatgatactctgaaggaacagcgtgtaagat- --gaagagaacccagagagacctga	4739
Db	780	ACCTGCACATGATCTCTTGACGAGACAGGTGAATTCGAGGGGCCCCAGAGGACCTTGA	839
QY	4740	gat 4742	
Db	840	GAT 842	
RESULT	9		
LOCUS	B1830310	770 bp	linear
DEFINITION	603073042P1 NIH_MGC_119	Homo sapiens	cdNA clone IMAGE:5164762 5',
ACCESSION	B1830310		
VERSION	B1830310.1	GI:15941860	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	NIH-MGC http://mgc.nci.nih.gov/ .		

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LIAW11409 row: d column: 11
High quality sequence stop: 760.

FEATURES
source
1. .770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5164762"
/clone_lib="NIH MGC.119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH MGC Library."
BASE COUNT 164 a 239 c 223 g 144 t
ORIGIN

Query Match 11.6%; Score 738; DB 10; Length 770;
Best Local Similarity 99.1%; Pred. No. 5.6e-123;
Matches 763; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

241 gaattcccaagctgactgtgactgtgctgtcgaagcccggaatgcgcgacacagag 300
|||||
1 GAATTCGCCAGCTGATGTGAGCTGTGCTGCTGCAGCCCGGGAATGCCGACGAG 60
301 ccgggagatcccaagagagaactgtgatgtccaggtgagggcggtgtgagatgat 360
|||||
61 CCCGGGATCCCPAGGATGAGAACTGATGCGCGAGGAGGCGCGGTGAGATGTAT 120
361 atttgagactgggtcatgtgtccacagaagatcagtaactgtgagtcagacacagccc 420
|||||
121 ATTGAGAGACTGGCTCATGTGTCACAGAAAGTATAGTACTGAGTGCAGATACAGCCCC 180
421 gtacacacagacacacagcgaggagcgacagaaaggcctcccgccaggtctttagag 480
|||||
181 GTACACACAGACACACAGCGGAGCGACAGAAAGGCTCTCCCGCCAGATCTTTAGCAG 240
481 gatgtctctgagacagagatgcggccctgagagactgaatgactccggcgctgctcg 540
|||||
241 GATGCTTCTGGAGACGAGAGNGTCCGCCCTGAGACTGGAATGACTCCCGGCTGCTCG 300
541 ggcctcccggaagacacccctcgaagcagtggtgctctagcatcttcgacctgagagac 600
|||||
301 GGCCTCCCGGAAGACACCCCTCGAAGCAGTGTGCTCTAGATCTTGAGACTGAGAAC 360
601 ctggagactgtactcatgtctgcctctctgtctagagcgggcgcccaagaatgtgagc 660
|||||
361 CTGGAGAGCTGACTCATGTCTGCTCTCTGTGCTAGAGCGGGGCGCCCGCAGAGATGTGAG 420
661 cgaggcaatgaacaccttcgaagcgagacacggccggccgctcaccctcaccgc 720
|||||
421 CGGCGAATGAACCTTCTGAGCGGACGCGGCCCGCGCTCTACCTTACCCG 480
721 gaacctgacgagatgaagacgctgtaagcgtgtgacgcgacacacaccccgagagac 780
|||||
481 GCACCTGACGAGAGATGAAGCGGTGAGAGCTGTATACCGCCCTAGACCCACCCCGCAGAC 540

0y 781 ttggacaaagagatctgtgtcaagtgtctgtcgtcaagltcagatltgaattgacc 840
|||||
Db 541 TTGGACAAAGAGATCTGTGTCAGAGTGTCTGCCCTCAAGTTGAGATTGAATTTAGAGCC 600
0y 841 atcttggagcttctgtgtcgtatgtgtcgggagagaaaagaatctccgagacttc 900
|||||
Db 601 ATCTTTGGGATCTTGGCTCTGTATGATGTGCGGAGAAAAAGAAATCTCGAAAACTTC 660
0y 901 tacttgacctgaactgtgactcc-atgaaggcgctcgttcggccatgagccaccacc 959
|||||
Db 661 TACTTGCACCTGAAGTGTGAGTCCATGAAGGGGCTGTCTCGGGCTCATAGGACCCACCC 720
0y 960 tggcatctcaacctgtgcccgtctgt-ccatctctctgtgacctacc 1008
|||||
Db 721 TGCCATCTCCACCTGTGCGCCGCTGTGCCCATCTTCTGTGAGATACCC 770

RESULT 10
A1198543/c 794 bp mRNA linear EST 10-NOV-1998
LOCUS
DEFINITION
g449f11.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1753389
3' similar to WP:146H5.4 CE04590 RAT TRG GENE PRODUCT ;, mRNA
sequence.
ACCESSION A1198543 GI:3751149
VERSION A1198543.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 794)
NCT-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
Unpublished (1997)
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCT-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html
Insert Length: 856 Std Error: 0.00
Seq primer: -40UP from G1bco
High quality sequence stop: 444.
Location/Qualifiers
1. .794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1753389"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGAGCGCCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES
source
1. .794
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1753389"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech Laboratories
, Inc., and primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAGTGGAGCGAGCGCCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 149 a 219 c 236 g 190 t
ORIGIN

Query Match 11.5%; Score 735; DB 9; Length 794;
Best Local Similarity 97.7%; Pred. No. 1.9e-122;
Matches 777; Conservative 0; Mismatches 15; Indels 3; Gaps 3;

|||||
Db 541 CGGCTACTTACCATGGCGGCTCTTACGAGCGCGTGAATGAGTCTTACAGAACTTCAT 600
Qy 5145 ccccatccttgagagccacccgtgactacaagaagctgagccgagtgacagggcaactgca 5204
Db 601 CCCCATCTCTGGAGGCCACCGTACTACAAAGACGTGGCCCGGTGACACGGCAAACTGCA 660
Qy 5205 ggaagccttaccacagatcatgacccagagttccgctgtagagcgctgtctcggaagcta 5264
Db 661 GGAGGCGTTTCACCAAGATCATGACACAGAGTTCCGGCTG---GAGCCCGTGTGGGACGT 717
Qy 5265 ttcccgctgagcttctacagcgccacctcgctgacctgagatgagcaagagt 5318
Db 718 ATTCCGCGTGGGCTTCTTACGGCGCCACTTCGTGGA-CTGGATGACAGAGATT 770
RESULT 12
AN027453/c 793 bp mRNA linear EST 15-SEP-1999
LOCUS Wt5403.x1 NCI-CGAP-GC6 Homo sapiens cDNA clone IMAGE:2515180 3'
DEFINITION similar to WP:F46H5.4 CE04590 RAT TRG GENE PRODUCT ; mRNA
sequence.
ACCESSION AM027453 GI:5886209
VERSION AM027453
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 793)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga903-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40up from Gibco
High quality sequence stop: 463.
FEATURES
source
Location/Qualifiers
1..793
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2515180"
/clone_lib="NCI-CGAP-GC6"
/tissue-type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI-CGAP-GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
1257096-1258631, 1459064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 153 a 229 c 227 g 183 t 1 others
ORIGIN
Query Match 11.3%; Score 720.4; DB 9; Length 793;
Best Local Similarity 94.7%; Pred. No. 8.3e-120;
Matches 745; Conservative 0; Mismatches 42; Indels 0; Gaps 0;
Qy 5572 acgacgttaccgagtgagcgagacagggagctgcccagcaacacagcgtgaag 5631

|||||
Db 787 AAGCTTATCAAGCGGATGGGAGAGACACGGAGTTCGCCAACCAACAGCTTAAG 728
Qy 5632 acgctgtccagacagacagacagccttccctacatacaagctcgatccgtgtgtccac 5691
Db 727 ACGATGCTTACAGAGCAAAAGCTTTCCTCCATCATCGAGCTCGCATCTGTGTCTCTC 668
Qy 5692 cggagagagacagtgatgacgacgttgagtgagtcacatgaagacatgacgaagacaa 5751
Db 667 GGTAAAGACAGCGTCTACGCGCCAGTGGAGGTGCATTCAGACATGCGAAGAAAGACA 608
Qy 5752 cggagacgtgaccttgcacacagcagagaccacacagatgctaatgatacatgagtg 5811
Db 607 CGGAGCTGGCGCTGTGCCACGAGACAGACACCAATGCTTAAGATGATACAGATGCTG 548
Qy 5812 cttaagagctctgttagagggccacagtgaaacagaggttccctgagatgagggcagaggttt 5871
Db 547 CTTCAAGGCTCTGTAGGACCCACCGGTGACAGAGGTCTCTCGAAGGTGGCCAGGTGTT 488
Qy 5872 ttagcagaagatcccggaagaccacagctctccgcatcacaaatgacgctctgc 5931
Db 487 TTAGCAGAGATCCCGAGAGACCCCAAGCTTTCGGCATCAACAATGCGGCTCTGC 428
Qy 5932 ttcaagagctctgtgaagaataatgtgagatgctgtgcgaaaaataaggccctgattgg 5991
Db 427 TTCAAGAGCTCTGCAAGAAATGTAGATGCTGCGGAAATTAAGCCCTGATTGGG 368
Qy 5992 ccgagacagaaagagttacacccgtgagctgagcgaactactgcccgtcgagaggt 6051
Db 367 CCGGACAGAAAGAGATATACACCGTAGCTGAGCGCAACTATCCCGCTGCGGAGGCT 308
Qy 6052 ctgcagccctgtgcttaccacagcgtgcccagctgtagtgacacccaccccgcttc 6111
Db 307 CTGCAGCCCCCTGTACCCACCGCTGCCCAATGTGATGGACCCACCCCGGCTTC 248
Qy 6112 aggaactccttgaaagacagcaagttccgaaagcagaacctgtgagcccaagagacaa 6171
Db 247 AAGAACTCCTTGAACAGAGCAAAAGTTTCGAAAGGACAGACCTGTAGCCCAAGAGACAA 188
Qy 6172 agctgtacctagaggaacacagaccccgagctcagctgtgtgtgtgtgtgtgtgtgt 6231
Db 187 AACTGTACTTACAGGAAACACACACCCCGGCTCAGCTGTGTCTCGAGGGAGAGCTG 128
Qy 6232 cccctgtgcccacatgagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6291
Db 127 CCTGTGTCGCCACTGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 68
Qy 6292 ctgtgtcccatctgt 6351
Db 67 CTGTGTCCCATCTGT 8
Qy 6352 taagcaa 6358
Db 7 TAAGCAA 1
RESULT 13
BF569035/c 709 bp mRNA linear EST 12-DEC-2000
LOCUS BF569035 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4300412 3',
DEFINITION mRNA sequence.
ACCESSION BF569035
VERSION BF569035 GI:11642415
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 709)
NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaabbs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plates: L16CM159 row: e column: 21
 High quality sequence start: 3
 High quality sequence stop: 707.

BASE COUNT	120 a	212 c	223 g	154 t
ORIGIN				

Query Match	10.7%	Score 683.4;	DB 10;	Length 709;
Best Local Similarity	99.6%	Pred. No. 3.8e-113;		
Matches 706; Conservative	0;	Mismatches 1;	Indels 2;	Gaps 2

[illegible]

Db 169 GGCAGACCTCTGACCCCAAGAGACCAAGGTACTAGAGGACAGACCCGGGCTT 110
 |||
 Qy 6204 cagctgtctgtgct-gcgaaaggagctgcacctggtgcacatggtctgtgggttgaca 626
 |||
 Db 109 CAGCTGTCTGTGCTGCGCAGGGAGTCTGCATGTCGCCACCTGGCTGTGGGTGACCA 50
 |||
 Qy 6263 cacttacttgggctgggacctgtgcacctgtgtccacatctgtgac 6311
 |||
 Db 49 CACTGTACTTGGGGCTGGGCCCTCTG-CCCTGTCTCCCAATCTGTGTC 2

RESULT	14
B1822534	
LOCUS	
DEFINITION	B1822534 745 bp mRNA linear EST 04-OCT-2001 603035242P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5176259 5' , mRNA sequence.
ACCESSION	B1822534
VERSION	B1822534.1 GI:15934084
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 745)
AUTHORS	NIH-MGC http://mgc.ncl.nih.gov/.
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. http://www.ncbi.nlm.nih.gov/mgc/staff/rstrausbergs.html .

BASE COUNT	147 a	238 c	215 g	145 t
ORIGIN				
Query Match		10.6%	Score 677.4	DB 10; Length 745;
Best Local Similarity		98.0%	Pred. No. 4.6e-112;	
Matches 728; Conservative		0;	Mismatches 11;	Indels 4; Gaps 4;

QY	3250	cgagctgcacactctctcagcgaagcccgagcccccaggtgacagcagctgtctgaact	3300
QY		1	CAAGAGTCCACCTTCTCCAGCCAGGCCCGGA-CCCAAGTGACGCGATGTTCAACG
Db			59
QY	3310	a-gtgaccattcggcagcgagcagcctctcctagctgagctctgtctgaagagacttgacact	3360
QY			
Db	60	AGGTGACCATTCGCGACGACGACACTTCTCACTGAGGCTCTCTCTATACGAGACTGGACT	119
QY	3369	gagccctgaacctgagctgagctgaagggcattctgtgtgcacaagaagccataagtgacct	3428

```
Db 120 GGGCCCTGAAACCTGAGGCTAAAGGGCATTCCTGTTGACACAGAAAGGCCATCAGTGTCTGT 179
QY 3429 gcaacagcgtctatgttggccatgatcacatgacccccgctacgagcgagcgcaactgtgaagc 3488
Db 180 GCACAGGCTCTATGTGGCCATGACACTGACCCCGGTACGCCGAGCCACTGTGAAGGC 239
QY 3489 tctgtggccgacactgtactgtccactgtcttgcattgtacagcgagataccttgcacgct 3548
Db 240 TCGTGTGGCCGACCTGACTGTCCACTGCTATGATGATGACAGGATACCTTGGCCACGGCT 299
QY 3549 gcatgactgtctgagagggccaggtcagcgtaacagactgtccctcaattgttgaactaga 3608
Db 300 GCATGACTTTCCTGAGAGGCCCCAGGTACAGCGGTAAAGCTGGCTCAATGCTTGACTAGA 359
QY 3609 cacagaagcgaaaggagacatgtgcggttacatcaaccctctgtggccattgagccattgc 3668
Db 360 CACAGAGGCGGAGGAGACATTCGGGTACCAATCAACCCCTGCTGTGSCATGGCCATTGC 419
QY 3669 tctgtggcccttgccctgtgctcccgggccacatctcccaaggggccaccaagcgcttc 3728
Db 420 TGGTGGCCCTTGAGCCCTGCTGCTCCGGGCGACATCTCCAGAGGCGCACACAGGCTTC 479
QY 3729 tctgagagcgtgtgcccctctctgtctgataagcgagacgttgcgtgtgtgtgtgtg 3788
Db 480 TCGCGAGGCTGTGCGCTCTCTCTGATGATCAAGCCGACCTTGTGCTGTGTGTGTGTG 539
QY 3789 ggtgtgtaaaacacacgagcgagcgtctctgacgagcgtgtggccactgtacactccc 3848
Db 540 GGTGCTGAAATAACACGAGCGGCGCTCCGACGCTGGGCCACTGACACTGACACTGCC 599
QY 3849 ccagctggagagcgtctgtgtgactgtgtacacttggctgtgctgtgagtaagaagg 3908
Db 600 CCGAGCTGGAGCGTGTGTGGA-TTGGTGTACCTTGTGCTGCTGCC-TTGAATGACAAAGG 657
QY 3909 gaaacagccttgaacgcatcaacagcctcacatcaaaaaatctctgtatataagagc 3968
Db 658 GAAAGAGGCTTGAAGCGCATCAACGCTCAATTCACAAATCTGTGATGAAAGGC 717
QY 3969 gacgtagaggaagcattctgtg 3991
Db 718 CGCGGTAGAGGAGCCATCTGGG 740

RESULT 15
LOCUS BG678867 857 bp mRNA linear EST 01-MAY-2001
DEFINITION BG678867 602624/67Ft_NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749689 5',
ACCESSION BG678867 mRNA sequence.
VERSION BG678867.1 GI:13910264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 857)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaab-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0603 row: e column: 18
High quality sequence start: 4
High quality sequence stop: 818.
Location/Qualifiers
```

```
source 1. 857
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749689"
/clone_1b="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (rtI phage-resistant)"
/note="Organ: Skin; Vector: pCMV-Sport6; Site: 1: NCI;
Site: 2: Sali; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT 209 a 254 c 247 g 147 t
ORIGIN

Query Match 10.6%; Score 673.8; DB 10; Length 857;
Best Local Similarity 92.2%; Pred. No. 2.1e-111;
Matches 790; Conservative 0; Mismatches 52; Indels 15; Gaps 7;

QY 5359 tcaacacgcttgagaggtctctacagagagatgttgcagagcgtctgtgatatc 5418
Db 1 TTACACCGGCTGGAGGAGTTCTACACGAGAGATTGGCCGACGCTGTGAGATTATC 60
QY 5419 aaagactctaacctgtgtggaacaagtcgaagcttgactacaaaaggcctacacagatc 5478
Db 61 AAAGACTCTTAACCTGTGTGACAACTCCAAAGCTTGTACTCAGAAAGGCTTACATCAATC 120
QY 5479 acgtatgtgaacgctacttgaactacacagagctcaagagcgggtgactacttgac 5538
Db 121 ACGTATGTGAAACCGTACTTGTGATTAACAGAGCTCAAGACCGGGTGACTACTTGTAC 180
QY 5539 cgaacatagtgcttgcacatctctgttctgcacgcccgttcaagccgagatggcgca 5588
Db 181 CGCAACTATGAGGCTTCCCAATTCCTGTGTGACGCGCTTCAACGCGGAGGGGCGCA 240
QY 5599 caccgagagctgcccggagcaacaacagcgttaagacgctgtcagacacgacacgacctc 5658
Db 241 CACGGGAGGCTGCCCGAGCAACACAGCAGTAAGACGCTGTGTACACCGACGACGCTTC 300
QY 5659 cccataatcaagactgcacatcgtgtgtgtgcacggagagagacggtgtgtgaagca 5718
Db 301 CCTTACATCAAGACTCCGATCCGATCCGATGTCACCGGAGGAGAGGAGGTCGACCCGATG 360
QY 5719 gaagtgcacatcgagagcagatgaagaagaacagagagcgtgtgtgtgtgtgtgtgtgt 5778
Db 361 GAGGTGCCCATCGAGGACATGACAGAAAGACAGGAGCTGTGTGTGTGTGTGTGTGTGTGT 420
QY 5779 gaccacacagatgctaaagatgctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5838
Db 421 GACCCACAGATGCTAAGATGCTACAGATGCTGCTTCAGGGCTGTGAGGGCCACCGTG 480
QY 5839 aaccaggtccctgtgaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5888
Db 481 AACCA-GGTCCCTCGAGAGGTGGCCAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 539
QY 5899 ctcttcgagatcaacaataatgtcgctgtgc-ttaagagacttgcgaagaatgtga 5957
Db 540 CTCTTCGGGATCAACAATAATGCGCTGTGCTTGAAGACTTGTGCAAGAAATGTGA 599
QY 5958 gcatgctgtcggaaataaagccctgtatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6016
Db 600 GATGCGCTCGGAAATAAAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 659
QY 6017 -agctgagcgcaactactgcgcctgtcggagagcgtctgtcaagccctgtctacc----- 6070
Db 660 AGCTTTGAAGGCAACTACTCGCGCTCGGGGAGGCTCTGTGAGCGCTTACCCCAAGC 719
QY 6071 -agcgctgtccacgtgtatgtgaaccacacccacccggtctagagacttcctt-gaaca 6127
Db 720 GCGCTGCCCAACTGATGAGGAGCCACCCCAAGCTTGTGAGAACTCTTGTGGAACA 779
QY 6128 gagcaag--tttcgaaagcagacgtctgtgaccacaaga--ccaagctgtactaga 6184
||||| || ||||| || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
```

Db 780 GAGCAAGGTTCCGGAAGGGAGACCTCTGAAGCCACAGAAGGCCAAAGCTGGACCTAGA 839
Qy 6185 gnaaccagcaccgagc 6201
Db 840 AGGAACCAAGCACGCGGC 856

Search completed: October 5, 2002, 04:27:31
Job time: 12634 sec

alignment_scores:

Quality: 10569.00 Length: 2047
Ratio: 5.163 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-736-968a-2 x AX173175 ..

Align seg 1/1 to: AX173175 from: 1 to: 6372

```
1 MetAlaIaSerGIuArGArGAlaPheAlaHisLysIleAsnArGrHVa 17
|||||
13 ATGGCTGCTCCGAGCCCGCGCTTGGCCACAAAGTCAACAGCGGT 62
|||||
17 LAlaIaGIuValArGIuLysGIuValSerArGIuArGIuArGIuSerProH 34
|||||
63 GGCCGAGAGGTGGGAACAGGTGTCCGGGAACGACAGTGTCCCGCC 112
|||||
34 LAsSerArGIuArGIuSerSerSerIeuGIuValProIeuThrGIuVal 50
|||||
113 ACTGCAGAGGCTGCAGAGCTCCCTGGGGGTCCACTGACTGAAATG 162
|||||
51 ValGIuProIeuAsPheGIuAsPValIleuIeuSerArGIuProAsPAl 67
|||||
163 GTCGAGCCCTGGACTTGAAGATGTACTTGTGAGCCGGCCACAGATGC 212
|||||
67 aGIuProGIuProIeuArGIuAsPleuValGIuPheProAlaAsPAsPleu 84
|||||
213 TGAGCCCGGGCCCTCAGGAGCTGGAGATTCACAGCTGATGACTTGG 262
|||||
84 LuIeuIeuIeuGIuProArGIuLysArGrHrHrGIuProGIuIlePro 100
|||||
263 AGCTGTGCTGCACGCCCGGGAAATGCCGACACAGGACCCGGGATCCC 312
|||||
101 LysAsPGIuLysIeuAsPAlaGIuValArGIuAlaValGIuMetGrHr 117
|||||
313 AAGGATGAATAAATGATGATGCCAGGTGAGGGCCGCGGTGAGATGTAT 362
|||||
117 eGIuAsPTrValIleValHisArGIuArGIuTrGIuTrIeuSerAlaIAT 134
|||||
363 TGAGGACTGGGTCAATGTCCACAAAGGTATCAGTACCTGAGTGCAGAT 412
|||||
134 YrSerProValThrThrAsPThrGIuArGIuArGIuLysGIuIleuPro 150
|||||
413 ACAGCCCGGTACACACAGACACACAGCGGAGACAGAGGCGCTCCCG 462
|||||
151 ArGIuValPheGIuIeuAsPAlaSerGIuAsPGIuArGIuSerGIuProGI 167
|||||
463 CGCCAGGCTTTGAGCAGATGCTTGGAGAGCAGAGTCCGCGCTGTA 512
|||||
167 uAsPSerAsPAsPArGIuArGIuSerGIuSerGIuSerProGIuAsPTrPro 184
|||||
513 GGACTCGAAATACTCCGGCGGTGGCTGGCGTCCGGGAAGACACCCCTC 562
|||||
184 YrSerSerGIuAlaSerSerIlePheAsPleuArGIuAsPleuAlaIAsP 200
|||||
563 GAACGAGTGGTGCCTTAGCATCTTGACCTGAGGAACCTGGAGCTGAC 612
|||||
201 SerIeuIeuProSerIeuIeuGIuArGIuAlaProGIuAsPAlaIAsPAr 217
|||||
613 TCATTGGCTGGCTCTGTGCTAGAGCGGGCGGCCCAAGAAATGTGGACCG 662
|||||
217 GaRqAsnGIuThrIeuArGIuArGIuHisArGIuProAlaIleuIeuThrL 234
|||||
663 GCGCAATGAACCTTGGAGGAGCAGCAGCGCCCGCGGCTGTCTACCC 712
|||||
234 euIuTrProAlaProAsPGIuAsPGIuAlaValGIuArGIuSerArGIuPro 250
|||||
713 TCTATCCCGGACCTGAGCAGAGATGAAGCCGTGGAACGCTGTACCCGCCA 762
|||||
251 GIuProProArGIuHisPheGIuGIuArGIuIleuValIuLysCysIeuSe 267
|||||
763 GAGCACCAGCGGAGCAGCTTTGGACAAGAGATTTGTGTCAGTGTCTGTC 812
```

```
267 rIeuLysPheGIuIleGIuIleGIuProIlePheGIuIleIeuAlaIeuT 284
|||||
813 GCTCAATTGCAAGTTGAATTAATGAGCCCACTTTGGGATCTTGGCTCTGT 862
|||||
284 YrAsPValArGIuLysLysLysIleSerGIuAsPheGrHrPheAsPleu 300
|||||
863 ATGATGTGCGGGAGAAAAAGAAATCTCGAGAACTTCTACTTCGACCTG 912
|||||
301 AsnSerAsPSerMetLysGIuIeuIeuArGIuAlaHisGIuThrHisProAl 317
|||||
913 AACTCGGACTCCATGAAGGGGCTGCTTGGGCTCATGAGCACCACCTTCC 962
|||||
317 aIleSerThrIeuAlaArGIuSerAlaIlePheSerValThrTrProSerP 334
|||||
963 CATCTCCACCTTGCCCGCTCTCCATCTTCTGTGTGACTTACCTCCATC 1012
|||||
334 roAsPilePheIeuValIleLysIeuGIuLysValIeuGIuIeuGIuIleu 350
|||||
1013 CTGACATCTTCTGTGATCATCAAGTTGGAGAAAGTCTTACGCAAGGGAC 1062
|||||
351 IleSerGIuCysGIuProTrMetValIeuLysGIuValAsPThrAl 367
|||||
1063 ATCAGTAGAGTGTGTGAGCTTCACTGATGTTGTAAGAAGTGGACACAGC 1112
|||||
367 aLysAsnLysGIuLysIeuGIuLysIeuArGIuAlaIleGIuIleuPheC 384
|||||
1113 CAAGACAAAGAGAGAGCTAGAGAAAGCTGCCTGGCGCGGACAGTCTT 1162
|||||
384 YrThrArGIuGIuArGIuArGIuArGIuArGIuArGIuArGIuArGIu 400
|||||
1163 GCACCGCCCTGGGGCGGCTGACCATGCTTCCCTCGGAGCGCGGCGCAC 1212
|||||
401 IeuAlaAsnIleValSerSerAlaGIuIeuAsPArGIuAsPAsPSe 417
|||||
1213 TTGGCCAAATCTGTGACAGCAGCTGGGAGCTGAGCAGGAGCTTGTACTC 1262
|||||
417 rGIuGIuGIuArGIuArGIuProAlaTrThrAsPArGIuArGIuArGIu 434
|||||
1263 GGAGGGGAGAGCGCGGCGCAGCTGGAGACAGCCCGCCCTCGGGGGCCCC 1312
|||||
434 LAsPArGIuAsPArGIuAsPAlaCysSerPheSerGIuPheArGIu 450
|||||
1313 AGGACCGGGGAGTGTGGGAGCAGACGCTTGCAGCTTCTGTGCTTCCGT 1362
|||||
451 ProAlaThrIeuThrValThrAsnPhePheLysGIuAlaGIuArGIu 467
|||||
1363 CCAGCCAGCTTAAGTGTACAAACTTCTTAAACAGAGAGCTGAGCGACT 1412
|||||
467 uSerAsPGIuAsPleuPheLysPheIeuAlaAsPMeTarGIuArGIuPro 484
|||||
1413 CAGTGACGAGAGCTCTTCAAGTTCTGTGCTGACATGAGAGGCGCCCTCG 1462
|||||
484 eRIeuIeuArGIuArGIuArGIuProValThrAlaGIuIleuLysIleAsP 500
|||||
1463 CCTGTGTGGCGGAGTACGCTGTGTACTGCCAGCTCAAGATGACAT 1512
|||||
501 SerProAlaProGIuAsnPProHisPheCysIeuSerProGIuIeuIeu 517
|||||
1513 TCTCGGCTCTGTAAATATCCCACTTGTGCTCTCCCTGAGAGCTGTTA 1562
|||||
517 sIleLysProTrProAsPProArGIuArGIuProTrHrLysGIuIleuGI 534
|||||
1563 TATCAACCTTACCCGAGCCAGGGGCGCGGCCACCAAGAGATTTCTGG 1612
|||||
534 LuPheProAlaArGIuValTrAlaProHisThrSerTrArGIuAsnIeu 550
|||||
1613 AGTTCCTCCGCGCGGAGGTATGTCCCTCATTCAGCTACAGGAACCTG 1662
|||||
551 LeuTrValTrProHisSerIeuAsnPheSerSerArGIuGIuIleuSer 567
|||||
1663 CTGTAGCTGTACCCGACAGCTTCACTTCAAGTCTGAGCAGCGCGCAGG 1712
```

```

567 1ArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProSerG 584
    |||||||
1713 GCGCAACCTTGCTGTGGCAGTGCATGATGACAGCGAGAGACCCAGGCC 1762
    |||||||
584 1nAlaLeuProVal1LePheGlySerSerCysSerGluPheThrArg 600
    |||||||
1763 AGGCTGTCCGGTCATCTTTGGCAAGTCCAGCTCAGTAATTACCCGC 1812
    |||||||
601 G1uAlaPheThrProValValTyrHisAsnLysSerProGluPheTyrG1 617
    |||||||
1813 GAGGCTTCACACCGGTGGTCTTACCATAACAAGTCCCGGAGTTCTAGCA 1862
    |||||||
617 uG1uPheLysLeuHisLeuProAlaCysValThrGluAsnHisThrLeu 634
    |||||||
1863 GGAGTTCAGCTGCATCTTCCAGCTGCGTGTACAGAGACCATCACTGCG 1912
    |||||||
634 eupheThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeu 650
    |||||||
1913 TGTTCACCTTCTACCATGTGTAGTGCAGCCCGCGGCGGACGTGCCCTG 1962
    |||||||
651 G1uThrProValGlyPheThrTyr1LeProLeuGlnHisGlyArgLe 667
    |||||||
1963 GAGACACCCCGGGGCTTACTTGGATCCACTGCTGCAGCAGCGGCCCT 2012
    |||||||
667 uArgThrGlyProPheCysLeuProAlaCysValAspGlnProProPro 684
    |||||||
2013 GAGAGACCGGCCCTTCTGTCTCCAGTGTGTGTGACACAGCCGCGGCCA 2062
    |||||||
684 eTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTyrVal 700
    |||||||
2063 GCTATTCGCTGCTCACACCCGATGTGGCGGTCCGGGCAATGGCTGGGTG 2112
    |||||||
701 AspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerVa 717
    |||||||
2113 GACGGTCAACAAGG3CGTTCAGTGTGAGTCAACAGCCGTCCTCTGT 2162
    |||||||
717 1HisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisVal 734
    |||||||
2163 GCACCCCGAGGACCCCTACCTGCGCAAAATCTTACCTGGTGGACGTCC 2212
    |||||||
734 euGlnGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 750
    |||||||
2213 TGGAGAGAGGAGCTTCCCATTCGGGCTCAAGGACACTGTGTGAGGAG 2262
    |||||||
751 GlyAsnValGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAl 767
    |||||||
2263 GGCACAGTGGAGCAGGAGCTGCGGGCCAGTCTTGACACTGGCGCTGGC 2312
    |||||||
767 aSerProGlnProLeuValAlaPheSerHisHisValLeuAspLysLeuV 784
    |||||||
2313 CAGCCCCGAAACCCCTTGTGGCTTCTCCACACAGTGTGAGCAAGCTCG 2362
    |||||||
784 a1ArgLeuVal1LeArgProPro1Le1LeSerGlyGln1LeValAsnLeu 800
    |||||||
2363 TGCCTGTGGTCATCAGGCCCCCGATCATGTGGCCGATGTGGAACCTG 2412
    |||||||
801 GlyArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisAr 817
    |||||||
2413 GGCCTGGAGAGCTTGAAGCAATGGCCCATGTAGTCAAGCTTGTTCACCG 2462
    |||||||
817 gSerLeuG1uAla1aGlnAspAlaArgGlyHisCysProGlnLeuAlaA 834
    |||||||
2463 GAGCTCGAGAGCAACCGAGATGCCCGGGTCACTGGCCACACACTGGCTG 2512
    |||||||
834 1aTyrVal1HisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro 850
    |||||||
2513 CCTACGTCACCTAGCTTTCGCTTCTGCTGCACTGAGCCACAGCTCCCG 2562
    |||||||
851 AspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgGlySe 867
    |||||||
2563 GATGGGGCCCTCCAGTACAGTGCAGGCTGCCACACTGGCCGCTGGCTC 2612
    |||||||
867 rg1yArgProAlaSerLeuTyrLeuAlaArgSerLysSer1LeSerSers 884
    |||||||
2613 TGGTGGCCCCGAGAGCTCTACTGGCGGTTCACAGACATCAGAGCA 2662
    |||||||
884 eArgAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
    |||||||
2663 GCACCCCTAGACCTCGCCGCGTGGCCCTGGGCTGTGTGATACAGAGTTCC 2712
    |||||||
901 Arg1LeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTyrVa 917
    |||||||
2713 CGCATCTGGCCAGCAAGCTGCTTCAGAGAGAGTGGCTCTCAGAGGGGT 2762
    |||||||
917 1ValSerSerSerAlaValArgGluAla1LeuGlnHisAlaThrPheP 934
    |||||||
2763 GGTACAGACAGTGGCCGTACGGAGGSCATCTCCAGACAGCTGGTGTCT 2812
    |||||||
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlnGly 950
    |||||||
2813 TCTTCAGAGTCATGAGTGAAGATAGGCGCTGACCTGCTGTGGCCAG 2862
    |||||||
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
    |||||||
2863 CGACTAGACACACCCCGCAAGCTGCGCTTCCCGGACGCTTCTGGACGA 2912
    |||||||
967 p1LeThrAlaLeuValGlySerValGlyLeuGluVal1LeThrArgValH 984
    |||||||
2913 CATCACTGCTTGTGTGGTCTGTGGGCTGGAGGTCAATACCCGCTGTCC 2962
    |||||||
984 1sLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
    |||||||
2963 ACAAGATGTGAGCTGGCGCGAGACACTCAAGCCAGCTGGCTTTCTTC 3012
    |||||||
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
    |||||||
3013 CTCAGTGAACCTTCTGTCCCTGTGTGACCGGGGCTTGTCTCACTGGT 3062
    |||||||
1017 1ArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerProAsp 1034
    |||||||
3063 CCGGGCCCACTACAAAGCAGGTGGCCAGCGGCTCCAGTGTCCCTTAATC 3112
    |||||||
1034 ro1Ala1aLeuLeuThrLeuArgMetGluPheThrArg1LeuGlySer 1050
    |||||||
3113 CAGCACCCCTGGTGAACCCCGGCATGTGATTCACCCGATCCTGTGGACG 3162
    |||||||
1051 HisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSerProPr 1067
    |||||||
3163 CAGSACACTAGTGAACCTCAACCTCCCTGCTGACCCCTGCACCTCC 3212
    |||||||
1067 oAlaSerProSerProSerValSerSerThrThrSerGlnSerSerThrp 1084
    |||||||
3213 AGCTTGGCCCTCCCTCTGTGTCTCCACACACTCCAGAGCTTCACCT 3262
    |||||||
1084 heSerSerGlnAlaProAspProLysValThrSerMetPheGluLeuSer 1100
    |||||||
3263 TCTCCAGCCAAAGCCCGGACCCCAAGGTGACCAAGCATGTTCACACTGAGT 3312
    |||||||
1101 GlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLe 1117
    |||||||
3313 GGACCATTCGGCAGACAGCATTCCTAGTGGGCTCCTGTGACGAGACT 3362
    |||||||
1117 uAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysL 1134
    |||||||
3363 GGCACCTGGCCCTCGAACTGAGGCTGAAGGGCATTCCTGTGGCAAGA 3412
    |||||||
1134 yAla1a1eSerAlaValHisSerLeuLeuCysGlyHisAspThrAspPro 1150
    |||||||
3413 AGGCGCATGCTGCTGTGACACAGCTGATATGTGGCAATGACACTACCCC 3462
    |||||||
1151 ArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyrLeuPr 1167
    |||||||
3463 CGCTAGCGCGGAGGCACATGTGAAGGCTGTGTGGCGGAGCTGTACTGCC 3512
    |||||||
1167 oLeuLeuSer1LeAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
    |||||||

```

3513 ACTGCTTCGATGCGAGGATACCTTGGCCAGCGCTGCATGACTTGGCTG 3562
1184 IUGLYProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
|||||
3563 AGGGCCCGAGGCTGACGAGCTGCAAGACTGCGCTCAATGCTTGACTGCAGACA 3612
1201 GUGUGlyGlnGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217
|||||
3613 GAAGCGCAAGGGGACATTCGGGCTACCATCAACCCCTCTGCGGCATGGC 3662
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGlnG 1234
|||||
3663 CATTCGCTGGTGGCCCCCTACGCCCTGGCTCCGGGCGACATCTCCAGG 3712
1234 LYPProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlnSerSer 1250
|||||
3713 GGCCACCAAGGCTTCTCGCGAGGCTGCGCTCTCTGCTGAGTCAAGC 3762
1251 ArgThrLeuLeuAlaCysValLeuTrpValLeuLysAsnThrGluProAl 1267
|||||
3763 CGGACCTTGCTGGCGCTGTGCTGTGGGTGCTGAAAAACACGAGCGCGC 3812
1267 aLeuLeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgL 1284
|||||
3813 GCTCTCGAGCGCTGGGCTGACTGACTGACCTCCCGAGCTGGAGCTC 3862
1284 euleuAspLeuLeuThrLeuCysLeuAlaAlaPheGluTrpLysGlyLys 1300
|||||
3863 TGTTCGACTTGTGCTGACTTGTGCTTGGCTGGCTTGTGAGTACAAAGGGAAA 3912
1301 LysAlaPheGlnArgIleAsnSerLeuThrPheLysLysSerLeuAspMe 1317
|||||
3913 AAGGCTTTGAACGCATCAACAGCCTCACATTCAAAAAATCTCTGCATAT 3962
1317 tLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgGln 1334
|||||
3963 GAAGGCGGCGCTAGAGAACCATCTGGGTACATCGGAGCTCGACAG 4012
1334 IuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350
|||||
4013 AATGCTTCGGGCAAGCTGTGAGAGAGACCCGTTTGGGAATCCAGAGAAC 4062
1351 ValArgTrpAlaGlySerValThrHisTrpLysGlnThrSerAspArgVal 1367
|||||
4063 GTGCGCTGGCGAAGAGCTGACACACTGGAAGAAACCTCAGACCGCGT 4112
1367 IAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384
|||||
4113 GGACAAAGCCAAAGCATGAATGGAACAGAGAGCCTTGGTGGAGAGGAAC 4162
1384 euAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIleIleVal 1400
|||||
4163 TGGCAACCGAGGCAAGCCTAGTGGTTTGGACACACTGAGATCATGCTG 4212
1401 GlnThrValMetLeuSerGluAlaArgGlnSerValLeuGlyAlaValLe 1417
|||||
4213 CAGAGGCTGATGCTTTCAGAAAGCCCGGAGAGCCTTGGGGGAGAGTCT 4262
1417 uLysValValLeuTrpSerLeuGlySerAlaGlnSerAlaLeuPheLeu 1434
|||||
4263 GAAGCTTGTGCTGTACAGCTGCGGACAGTCCCAAGTCCCGAGCTG 4312
1434 InHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeu 1450
|||||
4313 AGCATGGCTGGCCACCCAGAGGGCCCTTGTCCAAAGTCCCGAGCTG 4362
1451 LeuPheGlnGluAspThrGluLeuCysAlaAspLeuCysLeuAlaGlyLeu 1467
|||||
4363 CTGTTCCAGGAGGACACGAGCTGTGTCCGACCTGTGCTGAGAGCTCT 4412
1467 uArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerAlas 1484
|||||
4413 ACGACACTGTGTGACGCCCATTCAGACCATCCGACGCGACCGCGCT 4462
1484 eLeuThrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500
|||||
4463 CGCTGTACTGCTCATGTCGACAGAACTTCGAGATCGGCGCAACAATTGGC 4512
1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517
|||||
4513 CGTGTAGATGAGGTGACCATGTCCTCTCTGCTCCCTGGTGGGAGCAG 4562
1517 rGlnAsnPheSerGluGlnHisLeuArgArgSerLeuLysThrIleLeuT 1534
|||||
4563 GCAGAACTTCAGGAAGAGACACCTGGAGGTTGACTTCACTCAAAACACTCTCA 4612
1534 hTrpAlaGlnGluAspMetGlyLeuArgAspSerThrPheAlaGlnGln 1550
|||||
4613 CCTATGCTAGAGAGACATGGGGCTGGGAGACAGACCTTCGACAGAGCG 4662
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLys 1567
|||||
4663 GTCCAGGACCTGATCTTCAACCTGCACATGATCTGACGCGACGCGTGA 4712
1567 sMetLysGlnHisGlnGluAspProGluMetLeuIleAspLeuMetTyra 1584
|||||
4713 GATGAAGGAACACAGAGAGACCTGAGATGCTCATCGACCTCATGTACA 4762
1584 rGlnAlaArgGlyTrpGlnGlySerProAspLeuArgLeuThrTrpLeu 1600
|||||
4763 GAATTCGCCGGGCTACCAAGGGCTCACGAGCTTCGGCTGACCGTGTG 4812
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaL 1617
|||||
4813 CAGAACATGGCCGGGAGGACCGGGAGCTGGGCAACACAGCGGAGCGCC 4862
1617 aGlnCysMetValHisAlaAlaLeuValAlaGluTrpLeuAlaLeuL 1634
|||||
4863 CCAGTGCATGTGTGCACGCGCGCCCTCGTGGCTGAGTACTCGCCCTGC 4912
1634 euGlnAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIle 1650
|||||
4913 TCGAGAGCACGCGCCACCTGCTGGGCTGGCTTCTTCCAGAAATCC 4962
1651 SerSerAsnValLeuGlnGlnSerAlaIleSerAspAlaIleLeuSerPr 1667
|||||
4963 TCATCCACAGCTGCTAGAGAGTCCGCATCTCCGAGAACTCTGTGCGC 5012
1667 oAspGlnGlnGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuV 1684
|||||
5013 CGACGAGAGAGGCTTCTGCTCCGGGAAGCATTCATGAGCTGGGGCTGG 5062
1684 aGlyLeuLeuGlnGlnAlaAlaGlyTrpPheThrMetGlyGlyLeuTrp 1700
|||||
5063 TAGGTTGCTGGAAAGGACGCGGCTACTTCAACCTGGCGGCGCTTAC 5112
1701 GluAlaValAsnGluValTrpLysAsnLeuIleProIleLeuGlnAlaHis 1717
|||||
5113 GAGGCGGTGAATGAGGTCTTACAGAACTCATCCCATCTCGAAAGCCA 5162
1717 sArgAspTrpTrpLysLeuAlaAlaValaHisGlyLysLeuGlnGluAlaP 1734
|||||
5163 CCGTGACTACAAAGAACTGGCCGGGTGCACGGCAAACTGACGAGAGGCT 5212
1734 hThrLysIleMetHisGlnSerSerGlyTrpGluArgValaPheGlyThr 1750
|||||
5213 TCACCAAGATCAGTACACAGAGTTCGGGTGGAGGCGCTGTTCCGGAGG 5262
1751 TyrPheArgValaGlyPheTrpGlyAlaHisIlePheGlyAspLeuAspGlu 1767
|||||
5263 TATTTCGGGTGGCTTCTTACGCGCGCCCACTTCGGTACCTGAGATGAGCA 5312
1767 nGluPheValTrpLysGluProSerIleThrLysLeuAlaGluIleSerH 1784
|||||
5313 GGAAGTTTGTGTACAGAGGCATCGATCACGAAGCTGGAGAGATCTCAC 5362

1784 tsargleugluuiphetyrthrgluarphgclaspaspvalvalglu 1800
 |||||||
 5363 accggctggagggtttctacacggagatttggcgacacgctgctggag 5412
 |||||||
 1801 llellelaspaspeertyrprovalasplysseertyleuaspeertylly 1817
 |||||||
 5413 attatcaaaagactcttaccctgtggcgaagctcgaagcttgcacacaaa 5462
 |||||||
 1817 salatyrilleglnllethrttyrvalgluaprottyrphesprthrrtyglu 1834
 |||||||
 5463 ggctctac 5512
 |||||||
 1834 eulysasparvalthrttyrphesparvasttyrtylgluapthrrphe 1850
 |||||||
 5513 tcagagaccgggtgaccttaccctggcgaagcttgcacacacacacac 5562
 |||||||
 1851 leuphecythrprophenrproaspclaryalaahisgylgluapthr 1867
 |||||||
 5563 ctgttctgcacacccgttaccgctggatggcgccgacacacgggagactg 5612
 |||||||
 1867 ocluglnhislyslaryglvsthrleuleuserthrasphsialapheprot 1884
 |||||||
 5613 cgagcacaac 5662
 |||||||
 1884 yrillesthrarjlearyglvalcysahisargglugluuThrvalleuthr 1900
 |||||||
 5663 acatcaagactgcacatccctgtctgcccacggagagagagagctgctgac 5712
 |||||||
 1901 provalgluvalalailegluaspmetglnlyslsthrarjgluval 1917
 |||||||
 5713 ccagtgagagctgctcgcagagacacacacacacacacacacacacac 5762
 |||||||
 1917 aphealathrgluuinaspproproaspalalysmetleuamctval 1934
 |||||||
 5763 ctttgccac 5812
 |||||||
 5813 ttacagagctctgtaggccacacacacacacacacacacacacacac 5862
 |||||||
 1951 glnvalpheeluvalgluileprogluasprolyslleupheathghish 1967
 |||||||
 5863 cagggtgttttttagcagagatccccggaagaccccaagctcttccgcatca 5912
 |||||||
 1967 sasnllysluargcucysphelyasphecyslyslcysgsluaspa 1984
 |||||||
 5913 caacaaatgtggcgctctgcttcaagacttctgcagaaatgtgagatg 5962
 |||||||
 1984 lalauarlysasnllyslalaleuileglyproaspgluylgltuthis 2000
 |||||||
 5963 cgcttcggaaaaaafmaagcccttgattggcgacacacacacacacacac 6012
 |||||||
 2001 arggluugluargasnttyrcysargleuargglualaleuainprole 2017
 |||||||
 6013 cgtgacgtcgagacgcaactgctgctggcgagagcttgcacacacacac 6062
 |||||||
 2017 uluethrglnarjleuproglnleumetalaiprothrrprogluyleua 2034
 |||||||
 6063 gcttaccacagcgctgccccagctgattggcaccacacacacacacacac 6112
 |||||||
 2034 rganserleuasnllyslalaserphearjglysalaspplu 2047
 |||||||
 6113 ggaactccttgaaacagacagatttccgaaagcgacacactc 6153

seq_name: gb_AB037816

seq_documentation block:

LOCUS AB037816 4886 bp mRNA linear PRI 14-MAR-2000
 DEFINITION Homo sapiens mRNA for KIAA1395 protein, partial cds.
 ACCESSION AB037816
 VERSION AB037816.1 GI:7243170
 KEYWORDS Homo sapiens brain cDNA to mRNA, clone_lib:pbLuescriptII SK plus
 SOURCE

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE
 1 (sites)
 Nagase, T., Kikuno, R., Ishikawa, K. I., Hirose, M., and Ohara, O.
 Prediction of the coding sequences of unidentified human genes.
 XVI. The complete sequences of 150 new cDNA clones from brain which
 code for large proteins in vitro
 DNA Res. 7 (1), 65-73 (2000)
 JOURNAL
 MEDLINE
 20181126
 2 (bases 1 to 4886)
 Ohara, O., Nagase, T., and Kikuno, R.
 Direct Submission
 Submitted (31-JAN-2000) Osamu Ohara, Kazusa DNA Research Institute,
 Laboratory of DNA Technology, 1532-3 Yata, Kisarazu, Chiba
 292-0812, Japan (E-mail: cdna@info.kazusa.or.jp,
 URL: http://www.kazusa.or.jp/huge/, Tel: +81-438-52-3913,
 Fax: +81-438-52-3914)
 FEATURES
 source
 location/Qualifiers
 1..4886
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="h07927"
 /tissue="type="brain"
 /clone_lib="pbLuescriptII SK plus"
 1..4884
 /gene="KIAA1395"
 <1..>4884
 /gene="KIAA1395"
 /note="Start codon is not identified."
 /codon_start=1
 /product="KIAA1395 protein"
 /protein_id="BA092633.1"
 /db_xref="GI:7243171"

translation="DARTMASERRAFKINRTVAAYEVKOVRSKSGSHSRCS
 SLSGPILEVEYRPFDEVLISRPDAPGFLRVLPADLLELQPRCCTTEG
 IPKDKLQVRAVAEMTIEDWVTHRRYQYLSAAYSPVTDLQREKQGLPQVFEQ
 DAGDERSGPDSDNSRSGSPDDEPRSGASIFDLRNLAADSLPILERAPEQ
 VDRNEMTLRQHRPALLTLTPADEDEAVRCSRPEPRPFQRIYVLSLEAFET
 EIEPFGILAYVDEKKIISENFYDLNDSMGLRAHGTHTAISTLASATFST
 YSPDFIVITLKVLEVLQGGDISCECPGMYTKKKEKLEKLELAEROTRG
 RYRPFPAWTAHVHLANIVSAGDLRDSSEBERPANTDRRRRPPQASGDDACSF
 SGPRATITVYNTFKQAEKRLSDELEFKFMDMRPSLRLRPVIAQKIDISPA
 ENPHCSPELHLIKPYDPGRPLTKLEPAEVVAAPHTSYNLLYVPHSLNFS
 RGSVRNLAIVQYMTGEDPSQALPVIFGKSCSEFTEAFTPVYHNKSPFEYEEK
 LHPACTVENHLLFTTEYHVSQCPRGTALET PVGFWMIPILQHRRLRTGFCIPVY
 DQPPSYVLTPDVALPGMRWVDGKGVSELTAVASVHPDQYLLKPFYLVLEE
 GAFPRLDIVLSGNVDELRSLAALRIASPEPLVAFSHHTYDKYRVIYRPTIS
 GOIVNLGARGAEAAHVSYLYRSLLEAQRGCPDLAAVYHAFPLPTEFSLPDG
 APPYVQANATLARGSGRPASLYLARSKSISSNDPLAVAPQSVDEVSRILASKLHE
 ELALQWVSSAVREAILIQHAFPEFQMLNVSMALLAGQRLDTPKLRPGFRLDI
 TALQSVGLIEVITRVHKLDELAEHLNLSAFELSDLSLDVDRGFVSLVAHKOYAT
 RLOSSPNPALITLMEFTRLTCSHEHYVTLNLCPLSPASPSPSVTSOSSRF
 SSOAPDPKVTSMFELSGPFRQOHLAELLEALALEPARGAFLHKAASVHSL
 LCGHDPREYAEATKARVAELIPLUSTIARDTIPRIHDPABGQORSRLASMDST
 EGBEDIACTINPSTAMAIAGSPLBGRASISQPTASAGCALSESRTLLACVY
 VWLNTEPALLQRMATDLTPQLGRLLDLITPLCAAEYGAKAERINSITPKSLSD
 MKARLEAIICTIGARQEMVRSRERSRPFSGPENVWRKSVTHMKOTSDDVKTDEM
 EHEALVGENLATEASLVYDLTEIIVQVMLESEARISVGLAVLVYSLGSOALF
 LQELATQRAIVSKFPELLEFDELCAICLRLEKRSRISTIRHASASLYILMR
 ONFPIGNFARVKNQVMSLSLVGTQNSSEHRLRSKLTITLYAEEDGLDSTFA
 EOYVDLNFNLHITLDTYVKKKEHEDPEMLIDMTARIARIGSQSPDLRLTWLQNMAGK
 HAEIGNHAEACQVNAHA"

BASE COUNT 941 a 1605 c 1425 g 915 t

alignment_scores:

Quality: 8334.00 Length: 1624
 Ratio: 5.135 Gaps: 0
 Percent Similarity: 99.938 Percent Identity: 99.938

alignment_block:
 US-09-736-968A-2 x AB037816 ..

Align seg 1/1 to: AB037816 from: 1 to: 4886

1 MetAlaIaSerGIuArGArGAlaPheAlaHISLysILeAsnArGThrVa 17
 |||||
 13 ATGGCTGCTCCGAGCGCGCCCTTCGCGACAAGATCAACAGAGCGGT 62
 |||||
 17 LAlaAlaGIuValArgLysGIuValSerArGluArgSerGIySerProH 34
 |||||
 63 GCGCGAGAGGTGCGAAGCAGGTGTCGCGGAACGAGTGCTCCGCC 112
 |||||
 34 LsSerArGArGcYsSerSerSerLeuGIuValProLeuThrGIuVal 50
 |||||
 113 ACTCCACAGCGCGCTGCAGACGCTCCCTGGGGGTCCCACTGACTGAAGTT 162
 |||||
 51 ValGIuProLeuAspPheGIuAspValLeuLeuSerArGProProAspAl 67
 |||||
 163 GTCGAGCCCGCTGACTTTGAGAGATGTTCTGAGCGCGCCACAGATGC 212
 |||||
 67 aGIuProGIyProLeuArGAspLeuValGIuPheProAlaAspAspLeuG 84
 |||||
 213 TGAGCCCGCGGCCCTCAGGAGCCTGTGAATTCACAGTGAATGACTTGG 262
 |||||
 84 LuLeuLeuLeuGIuProArGIuLcysArGThrThrGIuProGIyLePro 100
 |||||
 263 AGCTGCTGCTGCAGCCCGGGAAATGCCGAGACAGAGCGCGCGGATGCC 312
 |||||
 101 LysAspGIuLysLeuAspAlaGIuValArgAlaAlaValGIuMetYrI 117
 |||||
 313 AAGGATGAAAAAATGATGCCAGTGAAGGCGCGGTGAGATGTATAT 362
 |||||
 117 eGIuAspTrpValIleValHISArGArGTyrGIuThrLeuSerAlaIaT 134
 |||||
 363 TGAGGACTGGGTCAATTGTCCAGAAAGATACATACCTAGTGCACAT 412
 |||||
 134 YrSerProValThrThrAspThrGIuArGIuArgGIuLysGIyLeuPro 150
 |||||
 413 ACAGCCCGCTGCACACAGACACAGCGGGAGCGACAGAAGGGCTCCCC 462
 |||||
 151 ArgGIuValPheGIuGIuAspAlaSerGIyAspGIuArgSerGIyProGI 167
 |||||
 463 CCGCAGGCTTTGAGCAGAGATGCTTGTGAGACAGAGGTCCGGCCCTGA 512
 |||||
 167 uAspSerAsnAspSerArGArGIySerGIySerProGIuAspThrPro 184
 |||||
 513 GGACTCGAATGACTCCCGGCTGGCTGGGCTCCCGGAAAGACACCCCTC 562
 |||||
 184 rGSerSerGIyAlaSerSerIlePheAspLeuArGAsnLeuAlaAlaAsp 200
 |||||
 563 GAAGCAGGTGCTCTAGCATCTTCGACCTGAGAAACCTGGCAGCTGAC 612
 |||||
 201 SerLeuLeuProSerLeuLeuGIuArgAlaAlaProGIuAspValaPar 217
 |||||
 613 TCATTGCTGCCCTCTCTGTAGAGCGGGCGGCCCAAGAAGTGTGGACGG 662
 |||||
 217 gArGAsnGIuThrLeuArGArGIuHISArGProProAlaLeuLeuThrL 234
 |||||
 663 GCGCAATAAACCTTCGACGCGACACGCGGCCCGCCCTGCTCACCC 712
 |||||
 234 euTYrProAlaProAspGIuAspGIuAlaValGIuArgcYsSerArGPro 250
 |||||
 713 TCTACCCCGGACCTGACGAGGATGAAGCCGTGGAACCTGTAGCGGCCCA 762
 |||||
 251 GIuProProArGArGIuHISPheGIyGIuArgIleLeuValLyscYsLeuS 267
 |||||
 763 GAGCACCCCGCGACACTTGTGCAAAAGGATCTTGGCAAGTGTGTGTC 812
 |||||
 267 rLeuLysPheGIuLLeuGIuLeuGIuProIlePheGIyIleLeuAlaLeuT 284
 |||||
 813 GCTCAAGTTCAGATTAATTAAGCCCATTTGGGATCTTGGCTCTGCT 862

284 YrAspValArgGIuLysLysLysIleSerGIuAsnPheTYrPheAspLeu 300
 |||||
 863 ATGATGTGGGGAGAAAAAAGATCTCGAGAACTTCTACTTGACCTG 912
 |||||
 301 AsnSerAspSerMetLysGIyLeuLeuArGAlaHISGIyThrHISProAl 317
 |||||
 913 AACTGGACCTCATGAAGGGGCTGCTGGGCTCAATGACCCACCCCTGC 962
 |||||
 317 aIleSerThrLeuAlaArGSerAlaIlePheSerValThrTYrProSerP 334
 |||||
 963 CATCTCCACCCCTGGGCCGCTCTGCATCTCTCTGTGACCTACCCCTAC 1012
 |||||
 334 roAspIlePheLeuValIleLysLeuGIuLysValLeuGIuGIuLysP 350
 |||||
 1013 CTGACATCTTCTCGTCAATGAGTGAAGAGTGCTTCACAGAGGGGAC 1062
 |||||
 351 lIleSerGIuLcysGIuProTYrMetValLeuLysGIuValaAspThrAl 367
 |||||
 1063 ATCAGTGAAGTCTGTGAGCCTTACATGGTGTGAAGAAGTGGACACAGC 1112
 |||||
 367 aLysAsnLysGIuLysLeuGIuLysLeuArgLeuAlaIaGIuGIuPheC 384
 |||||
 1113 CAAGAACAAGAGAGCTAGAGAACTGCGCTGGCGCGCGAGCACTTCT 1162
 |||||
 384 YrThrArgLeuGIyArgTYrArGMetProPheAlaTrpThrAlaValHIS 400
 |||||
 1163 GCACCGCGCTGGCGCGCTACCCATGCCCTTGCCTGGACGCGCCCTGCAC 1212
 |||||
 401 LeuAlaAsnIleValSerSerAlaGIyGIuLeuAspArGAspSerAsp 417
 |||||
 1213 TTGGCCAAATCGTAGAGCGCGCTGGGAGCTGGACCGGACTCTGACTC 1262
 |||||
 417 rGIuGIuLysArGArGProAlaTrpThrAspArGArGArgGIyProG 434
 |||||
 1263 GGAGGGCGAGCGCGCGCGAGCTCGAGACAGACCGCGCGTGGGGGCC 1312
 |||||
 434 lAspArGAlaSerSerGIyAspAspAlaLysSerPheSerGIyPheArg 450
 |||||
 1313 AGGACCGGGCAGTGTGTGGGAGCAGCGCTGCAGCTTCTGTGGCTTCCGT 1362
 |||||
 451 ProAlaThrLeuThrValThrAsnPhePheLysGIuAlaGIuArgLe 467
 |||||
 1363 CCAGCCACCTTAACGTCTCAAAATCTTTAAGCAGAGGCTGAGCAGCT 1412
 |||||
 467 uSerAspGIuAspLeuPheLysPheLeuAlaAspMetArGArGProSers 484
 |||||
 1413 CAGTACGAGGACCTTTCAGTCTCGGCTGACATGAGGCCCGCTGCT 1462
 |||||
 484 eTLeuLeuArGArGLeuArGProValThrAlaGIuLeuLysIleAspIle 500
 |||||
 1463 CCTGCTGGGGGACTACGTCTGTGACTGCCAGCTCAAGATCCACAT 1512
 |||||
 501 SerProAlaProGIuAsnProHISPheCysLeuSerProGIuLeuLeuH 517
 |||||
 1513 TCTCGGCTCTGTGAANAATCCCACTTCTGCTCTCCCTGAGACTCTTCA 1562
 |||||
 517 sIleLysProTYrProAspProArgGIyArgProThrLysGIuLLeuG 534
 |||||
 1563 TATCAAGCCCTACCCCGACCCAGGGCGCGGCCCAAGAGGATTTGG 1612
 |||||
 534 LuPheProAlaArgGIuValTYrAlaProHISThrSerTYrArgAsnLeu 550
 |||||
 1613 AGTTCGCCCGCGGAGTCTATGCCCCCATACCAAGCTACAGGAACCTG 1662
 |||||
 551 LeuTYrValTYrProHISerLeuAsnPheSerSerArGGIuLysSerVa 567
 |||||
 1663 CTGTAGTGTACCCGACAGCCCTCAAACTTCACAGCGCGCCAGGCTCGCT 1712
 |||||
 567 lArGAsnLeuAlaValArgValGIuTYrMetThrGIyGIuAspProSerG 584
 |||||
 1713 GCGCAACCTTGTGCTGCGAGTGCAGTACATGACAGCGGAGGCCACCCAGCC 1762
 |||||
 584 lAlaLeuProValIlePheGIyLysSerSerCysSerGIuPheThrArg 600


```

|||||
1763 AGGCTTGGCGGTCATCTTTGGCAAGTCCAGCTGCAGTGAATTTACCCG 1812
601 G|U|A|A|P|H|e|T|h|P|r|o|V|a|I|T|h|H|S|u|N|S|e|S|e|r|P|r|o|G|I|u|P|h|e|T|h|Y|G|I| 617
|||||
1813 GAGGCGTTACACACCGTGCTGTACCATAAACAAGTCCCCGAGTTCTACGA 1862
617 uG|U|P|h|e|I|S|L|e|u|H|I|S|L|e|u|P|r|o|A|I|C|S|V|a|I|P|h|I|u|S|n|H|I|S|L|e|u| 634
|||||
1863 GGAGTTCAAGCTGCATCTTCCAGCTGGCTGCACAGAAACCATCACTGCG 1912
634 e|U|P|h|e|T|h|P|h|e|T|h|H|S|V|a|I|S|e|r|C|S|G|I|n|P|r|o|A|r|P|r|o|G|I|T|h|H|a|L|e|u| 650
|||||
1913 TGTTCACCTTCTACCATGTCAAGTGCAGCCCGCGGCGACCTGCTG 1962
651 G|I|u|T|h|P|r|o|V|a|I|G|I|P|h|e|T|h|T|r|I|P|r|o|L|e|u|G|I|n|H|I|S|G|I|A|r|G|L|e| 667
|||||
1963 GAGACACCGGTGGCTTACTTGATCCACAGTGCAGACAGGCGGCT 2012
667 u|A|r|G|h|I|G|I|P|r|o|P|h|e|C|S|L|e|u|P|r|o|V|a|I|S|e|r|V|a|I|S|p|I|n|P|r|o|P|r|o|S| 684
|||||
2013 GAGGACCGGCGCCCTTCTCTCCAGTGTCTGTGACCAAGCGCGGCCCA 2062
684 e|T|h|S|e|r|V|a|I|L|e|u|T|h|P|r|o|A|S|P|V|a|I|A|L|e|u|P|r|o|G|I|M|e|T|A|r|G|h|I|P|V|a|I| 700
|||||
2063 GGTATTCGTGCTCACACCCGATGGCGCTTCCGGGCAATGGCTGGGTG 2112
701 A|s|P|G|I|H|I|S|L|S|G|I|V|a|I|P|h|e|S|e|r|V|a|I|u|L|e|u|T|h|H|a|V|a|I|S|e|r|S|e|r|V|a| 717
|||||
2113 GACGGTCAACAAGGCGGTGTCAAGTGTGAGCTCAGCGCGCTGTCTCTGT 2162
717 I|H|I|S|P|r|o|G|I|n|A|S|P|r|o|I|T|h|L|e|u|A|S|P|L|S|P|h|e|P|h|e|r|h|I|e|u|V|a|I|H|S|V|a|I|L| 734
|||||
2163 GCACCCCAAGACCCCTTACCTGGAGAAATTTCTTCACTCGTGTGACGCTC 2212
734 e|U|G|I|u|G|I|Y|A|I|P|h|e|r|o|P|h|e|r|g|L|e|u|I|S|A|S|P|h|I|V|a|I|L|e|u|S|e|r|G|I|u| 750
|||||
2213 TGGAGAGAGGAGCCCTCCCATTCGGGCTCAAGACACTGTCTGAGCGAG 2262
751 G|I|Y|A|S|N|V|a|I|G|I|u|G|I|n|L|e|u|A|r|G|A|S|e|r|I|e|u|A|a|I|a|L|e|u|A|r|G|L|e|u|A|I| 767
|||||
2263 GGCAACGTGGAGCAGGAGCTGGGGCCAGTCTTCAGACACTGCGCCTGGC 2312
767 a|S|e|r|P|r|o|G|I|u|P|r|o|L|e|u|V|a|I|A|P|h|e|S|e|r|H|I|S|V|a|I|L|e|u|A|S|P|L|S|I|e|u|V| 784
|||||
2313 CAGCCCCAACCCTTGCGCTTCTCCACACAGCTGTGACAAAGCTCG 2362
784 a|A|r|G|L|e|u|V|a|I|I|e|r|P|r|o|P|h|e|r|I|e|S|e|r|G|I|u|I|L|e|V|a|I|S|N|L|e|u| 800
|||||
2363 TGGCTGTGGTCATCAGGCCCGCATCATCAGTGGCCAGATTGTGAACCTG 2412
801 G|I|Y|A|r|G|G|I|Y|A|I|P|h|e|G|I|u|A|I|e|T|h|A|H|I|S|V|a|I|S|e|r|I|e|u|V|a|I|H|I|S|A|r| 817
|||||
2413 GGCGGTGGAGCCTTGAAGCAATGGCCCATGTAGTCAAGCCTTGTTCACCG 2462
817 g|S|e|r|L|e|u|G|I|u|A|I|a|G|I|n|S|P|A|I|A|r|G|I|Y|H|I|S|C|S|P|r|o|G|I|n|L|e|u|A|I|A| 834
|||||
2463 GAGCCTGGAGGACCCAGATGCGCCGCGGTCATGCCCCACAGCTGGCTG 2512
834 G|A|G|C|T|G|G|A|G|G|C|C|A|G|C|C|A|G|A|T|G|C|C|G|C|G|C|A|C|T|G|G|C|T|G|G|C|T|G| 2562
2513 CCTACGTCCATCAGCCTTTCGCTTCTGACACTGAGCCAGCGCTCCCG 2562
851 A|S|P|G|I|Y|A|I|P|r|o|P|r|o|V|a|I|T|h|V|a|I|G|I|n|A|I|a|L|e|u|H|I|e|r|L|e|u|A|a|a|r|G|I|S|e| 867
|||||
2563 GATGGGGCGCTCCAGTGCAGTGCAGGCTGGCCCACTGGCGCTGGCTC 2612
867 r|G|I|Y|A|r|P|r|o|A|S|e|r|L|e|u|T|h|L|e|u|A|A|r|G|S|e|r|I|S|e|r|I|S|e|r|S| 884
|||||
2613 TGTGCGCGCCCAAGCCTTACCTGCGGGCGCTTCCAAAGCATCAGCGCA 2662
884 e|r|A|S|N|P|r|o|A|S|P|L|e|u|A|I|a|V|a|I|a|P|r|o|G|I|S|e|r|V|a|I|a|S|P|G|I|u|V|a|I|S|e|r| 900
|||||

```

```

2663 GCAACCCGTGACCTCGCCGTGGCCCTGTGCTGTGATGACGAGTTTCC 2712
901 A|r|G|I|L|e|u|A|I|S|e|r|I|S|L|e|u|H|I|S|G|I|u|L|e|u|A|I|a|L|e|u|I|n|T|r|P|V|a| 917
|||||
2713 CGCATCCTGGCGACCAAGCTGTTCACAGAGAGTGTGCTGTGACGTGGGT 2762
917 I|V|a|I|S|e|r|S|e|r|I|a|V|a|I|A|r|G|I|u|A|I|a|I|L|e|u|G|I|n|H|I|S|A|I|a|I|T|r|P|h|e| 934
|||||
2763 GGTAGCAGCAGTCCCTACCGAGGCCATCTCCAGACAGCCTGGTTCT 2812
934 h|e|r|I|G|I|n|L|e|u|M|e|T|V|a|I|L|S|e|r|M|e|T|V|a|I|a|L|e|u|H|I|S|L|e|u|L|e|u|G|I|n| 950
|||||
2813 TCTTCCAGTCAATGTTGAAGATGTGGCTGCACCTGCTGCTGGCCAG 2862
951 A|r|G|L|e|u|S|P|h|I|P|r|o|A|r|G|I|S|L|e|u|A|r|G|P|h|e|r|P|r|o|G|I|Y|A|r|G|h|e|u|S|P|S| 967
|||||
2863 CGACTAACAACACCCCGCAAGCTGCGCTTCCCGAGCGCTTCTGAGACA 2912
967 P|I|e|r|h|I|a|L|e|u|V|a|I|G|I|S|e|r|V|a|I|G|I|L|e|u|G|I|u|V|a|I|I|e|T|h|A|r|G|V|a|I|H| 984
|||||
2913 CATCAGTCCCTTGGTGGCTCTGTGGGCTGGAGGTCAATCACCCGTGTC 2962
984 I|S|L|S|A|S|P|V|a|I|G|I|L|e|u|A|I|G|I|H|I|S|L|e|u|S|N|A|I|S|e|r|I|e|u|A|I|a|P|h|e| 1000
|||||
2963 ACAAGGATGTGGAGCTGGCCGAGCACCTCAACGCCAGCGCTGTCTTCTC 3012
1001 L|e|u|S|e|r|S|P|L|e|u|S|e|r|I|e|u|V|a|I|A|S|P|A|r|G|I|Y|P|h|V|a|I|P|h|e|r|I|e|u|V|a|I| 1017
|||||
3013 CTCAGTACCTTCTGTGCCCTGTGTGGACCGGGCTTTGTCTTCAGCTGGT 3062
1017 I|A|r|G|A|H|I|S|T|h|L|S|G|I|n|V|a|I|a|T|h|A|r|G|L|e|u|G|I|n|S|e|r|S|e|r|P|r|o|A|S|P| 1034
|||||
3063 CCGGGCCCACTACAGAGAGGTGGCCAGCGCGCTCCAGTGTCTCCCAATC 3112
1034 r|o|A|I|a|I|a|L|e|u|T|h|L|e|u|A|r|G|M|e|r|G|I|u|P|h|e|r|h|I|e|u|Y|C|S|S|e|r| 1050
|||||
3113 CAGCAGCCCTGCTGACCTGCGCATGGAATTCACCGCATCTCTGTACAG 3162
1051 H|I|S|G|I|u|I|S|T|h|Y|V|a|I|T|h|L|e|u|A|S|N|L|e|u|P|r|o|C|Y|S|C|P|r|o|L|e|u|S|e|r|P|r|o| 1067
|||||
3163 CACGAGCACTAGTGAACCTCAACCTCCCTGCTGCCCTGCACCTCC 3212
1067 c|A|I|S|e|r|P|r|o|S|e|r|P|r|o|S|e|r|V|a|I|S|e|r|T|h|T|h|S|e|r|G|I|n|S|e|r|T|h|P| 1084
|||||
3213 AGCCTGCGCCCTCCCTGTGTCTTCACACACCTCCAGAGCTCCACCT 3262
1084 h|e|r|S|e|r|G|I|n|A|I|P|r|o|A|S|P|r|o|L|S|V|a|I|T|h|S|e|r|M|e|r|h|e|G|I|u|L|e|u|S|e|r| 1100
|||||
3263 TCTCCAGCCAGACCCCGGAGCCCAAGGTGACCAAGCTTTCGAACTGAGT 3312
1101 T|C|T|C|A|G|C|C|A|G|C|C|C|G|A|C|C|C|A|A|G|T|G|A|C|A|G|C|A|T|T|T|C|G|A|C|T|G|A|C|T| 3312
|||||
3313 GGACCATTCGCGCAGCAGCACTTCTAGCTGGGCTCTCTGCGGAGCT 3362
1117 u|A|I|a|L|e|u|A|I|a|L|e|u|I|n|P|r|o|G|I|u|A|I|a|G|I|u|G|I|Y|A|I|P|h|e|L|e|u|H|I|S|L|S| 1134
|||||
3363 GGCACGTGGCCCTCGAACCTGAGGCTGAAGGGCATTCCTGTTGCACAGA 3412
1134 y|S|A|I|a|I|S|e|r|I|a|V|a|I|H|I|S|S|e|r|L|e|u|C|Y|S|G|I|Y|H|I|S|P|h|I|S|P|r|o| 1150
|||||
3413 AGGCATCACTAGTGTGACACAGCTCTATGTGGCCATCACTGACATGACCC 3462
1151 A|r|G|T|h|A|I|G|I|u|A|I|a|T|h|V|a|I|Y|S|A|I|a|r|G|V|a|I|a|G|I|u|L|e|u|T|h|L|e|u|P|r| 1167
|||||
3463 CGCTACGCGGAGGCCACTGTGAAGCTCTGTGGCGGAGCTGTACTGTGC 3512
1167 o|L|e|u|S|e|r|I|e|a|I|a|r|G|S|P|h|L|e|u|P|r|o|r|g|L|e|u|H|I|S|A|P|h|e|I|a|I|a|G| 1184
|||||
3513 ACTGCTTGATTGACAGGGATACCTTGCCACGCGTGCATGACTTTGGCTG 3562
1184 I|u|G|I|P|r|o|G|I|u|A|r|G|S|e|r|A|r|G|L|e|u|A|I|S|e|r|M|e|T|L|e|u|A|S|P|S|e|r|A|S|P|h| 1200
|||||
3563 AGGGCCAGGTGAGCGGTCAAGACTGGCCTCAATGCTTGACTCAAGACACA 3612

```

1201 GluGlyGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAla 1217
|||||
3613 GAAGGCGAAGGGGACATTCGGGTACCATCAACCCCTCTGTGGCCATGCG 3662
1217 aileAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGlnG 1234
|||||
3663 CATTGCTGGTGGCCCCCTAGCCCCCTGGCTCCGGGGCCAGCATCTCCAGG 3712
1234 1yProPrrhAlaSerArgAlaGlyCysAlaLeuSerAlaGlySerSer 1250
|||||
3713 GGCACCAACAGGCTTCGCGCAGGCTGTGCCCTCTGCTGAGTCAAGC 3762
1251 ArgThrLeuLeuAlaCysValIleLeuThrValLeuLeuAsnThrGluProAl 1267
|||||
3763 CGAGCCTTGCTGGCTGTGTGTGGTGTCTGAAAAACACCGAGCGCGC 3812
1267 aLeuLeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgL 1284
|||||
3813 GCTCTTCACAGCGCTGGGCCACTGACCTGACACCTCCCGACGCTGGAGCTC 3862
1284 euleuAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrGlyGlyLys 1300
|||||
3863 TGTGGACTTGTGTACCTTGGCTGTGGCTGTGAGTACAGGGGAGAA 3912
1301 LysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspPhe 1317
|||||
3913 AAGGCTTTAAACGCATCAACACAGCTCACAATCAAAAAATCTGTGATAT 3962
1317 tLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGlyAlaArgGlnG 1334
|||||
3963 GAAGCGCGCGCTAGAGGAAAGCCATTCTGGGTACCATCGAGCTCGACAAG 4012
1334 LuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350
|||||
4013 AAATGGTTCCGCGAAGTCGTAGAGGAGCCCGTTGGGAATCCAGAGAAC 4062
1351 ValArgTrpArgLysSerValThrHisTrpLysGlnThrSerAspArgVal 1367
|||||
4063 GTGGCGTGGGGGAAGAGCGTCACACACTGAGAGAAACCTCAGACCGCGT 4112
1367 LAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384
|||||
4113 GGACAAACACCAAGCATGAATGAAACAGGCGCTGTGGAGGGGAAACC 4162
1384 euAlaThrGlnAlaSerLeuValValLeuAspThrLeuGlnIleIleVal 1400
|||||
4163 TGGCAACCGAAGCGCTAGTGTGTGGACACACTGGAGATCATCTGCG 4212
1401 GlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLe 1417
|||||
4213 CAGACGGTGTGCTTCAGAAAGCCCGGAGAGCGCTTGGGGCAGAGTCT 4262
1417 uLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeu 1434
|||||
4263 GAAGCTTGTGTGTACAGCCCTGGGCACTGCCACAGTGGCCCTTCTTGGC 4312
1434 LnhISGLyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeu 1450
|||||
4313 AGCATGGCGCTGGCCACCAAGGGCCCTGTGTCCAAAGTTCCGGAGCTG 4362
1451 LeuPheGluGlnAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLe 1467
|||||
4363 CTGTTCCGAGGAGACAGGAGCTGTGTCCGACCTGTGCTGAGGCTCTCT 4412
1467 uArgHisCysGlySerArgLysSerThrIleArgThrHisAlaSerLys 1484
|||||
4413 AGCAACACTGTGGCAGCGCATCAGCACCATCCGACACCGCCAGCCGCT 4462
1484 eLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500
|||||
4463 CGCGTACCTGCTCATGGAGACGAAGACTTCGAGATCGGCCACAACTTGGC 4512

1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517
|||||
4513 CGTGTGAAGATGCAGAGTACCATGTCTCTCTGCTCCCTGTGGGGAGAC 4562
1517 rGlnAsnPheSerGluGlnHisLeuArgArgSerLeuLysThrIleLeuT 1534
|||||
4563 GCAGAACTTCAGTGAAGAGACACTGCGACGTTCACTCAAAAACATCTCTCA 4612
1534 hTyrAlaGluGlnAspMetGlyLeuArgAspSerThrPheAlaGluGln 1550
|||||
4613 CCTATGCTGAGAGACATGGGGCTGCGGACACACCTTGTGCAAGACAG 4662
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrVally 1567
|||||
4663 GTCCAGGACCTGATGTTCAACCTGACATGATCTCGAGCGACAGGTGAA 4712
1567 sMetLysGlnHisGlnGlnAspProGlnMetLeuIleAspLeuMetTyr 1584
|||||
4713 GATGAAGGAAACACAGAGAGACCTGAGATGCTCATCCACTCATGTACA 4762
1584 rGileAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTyrLeu 1600
|||||
4763 GAATGGCCCGGGCTACCAAGGCTTACCGGACCTTGGCTGACTGCTG 4812
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGlnLysHisAlaGluAla 1617
|||||
4813 CAGACATGCGCGGGAAGCAGCGGAGCTGGGCAACACCGCGAGCGCGC 4862
1617 gLncGlyMetValHisAlaAla 1624
|||||
4863 CCAGTGCATGTGTGCACGCGGCC 4884
seq_name: gb_pat:AX172874
seq_documentation_block:
LOCUS AX172874 6828 bp DNA linear PAT 03-JUL-2001
DEFINITION Sequence 1 from Patent WO0142297.
ACCESSION AX172874
VERSION AX172874.1 GI:14597911
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 6828)
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.
TITLE Clasp-3 transmembrane protein
JOURNAL Patent: WO 0142297-A.1 14-JUN-2001;
Arbor Vita Corporation (US)
FEATURES
Source
Location/Qualifiers
1..6828
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="full-length human CLASP-3 cDNA"
22..6294
/note="human CLASP-3"
/codon_start=1
/protein_id="C043726.1"
/db_xref="GI:14597912"
/translation="MAERAFAGKISRTVAEVRKQISQYSGSPQLLNINIVNIS
HHTTPTLEAVDVEDLEDYLTIRPLAVDSGLDLIEPPDIEVYSPRDCITLYSA
VPESEMDPHVRCIRSYEDMAIVIRKHYKLGTFNPMTLDKROKROGLPQVES
DEAPDNGSYDDDDLLKRKMSIDTTPRGSMCSITPDILKNSLPDALPLDPTPHEE
IDKQNDQKRSNHRKELFALHPSFDEDEPLETSLVSDPIREHFGQLVKCYSIKETI
EIEPIFASLALYVKEKKISENFYDLNSEDQMKGLRPHVPPAALTTLARSAIFET
YSPQDVLVLIKLEKVLQOQDICECAPYMFKEADATKNKLEKLSQADQFCQGLG
KYRMPFMTAIHLMNIYSSAGSLRDSSTEYESTGRKGSERMSIYGRSLERT
TSGDQACNTLSFRPATLTVTNPFKQSGDRUSDLYLKKPLADMRPSSVLRRLPIAAO
LKIDISAPRNPPIHCLTPELLOVKLYLPDSDVRRTRLELRPADVYVPPVTRKLLYI
YPSIDANRQGSARNITTVKQFMYGEDPSNAPVIFGSSCESEKATVAVYHNR
SPDFHEIKVKVLPATLTDHNLHLETFYHNSQDQKQNTPLETGVGYTMTPLQNGRLKT
GQFLPVSLEKPPQAYSLSPEVPLPGMKVVDHMKGVFNVEYVAVSIITQDPYLDKF
FALYNALDEHLFPRIQDMRIEMENILELKSISALNSQLEPVPVVRFLHLLDLIL


```

1369  ....:|||||.....:|||||
1369 ACACGAGTGGAGTGTGCTTGTACTACAGCTTTCGACCACCTAC 1418
453 rLeuThrValThrAsnPhenylsGlnGluAlaGluArgLeuSerAspG 470
|||||.....:|||||
1419 TCACAGCTGACCAAAATTTTAAAGCAGAGAGACCGCTTAAGTGATG 1468
470 LuAspLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeu 486
|||||.....:|||||
1469 AAGATCTCTCAAAATCTCTGCTGATGTGAGAAAGCCATCTCTGCTTA 1518
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
|||||.....:|||||
1519 CGGGAGCTAGAGCTATTACAGCTCAGCTCAGATGACATTTCTCCCGC 1568
503 aProGluAsnProHisPheCysLeuSerProGluLeuLeuHisIleLysP 520
|||||.....:|||||
1569 ACCGTAAATCCCATTTATTCCTTAAGCTCCGAGCTGCTTCAAGTGAAGC 1618
520 rofYrProAspProArgGlyArgProThrLysGluIleLeuGluPhePro 536
|||||.....:|||||
1619 TTTACCTGTGACAGTAGAGTTAGACCTCCAGAGAAATCTTAGAGTTCCC 1668
537 AlaArgGluValTyrAlaProHisThrSerTyrArgAsnLeuLeuTyrVa 553
|||||.....:|||||
1669 GCAAGGAGTGTATTGTCCAAACACTACTTACAGAAATCTCTCTACAT 1718
553 lTyrProHisSerLeuAsnPhenSerSerArgGlnGlySerValArgAsnL 570
|||||.....:|||||
1719 ATACCTCCAGAGCTTATATTTTCCCAATCGTCAAGGTTCTGCTGAAATA 1768
570 euAlaValArgValGlnTyrMetThrGlyLysAspProSerGlnAlaLeu 586
|||||.....:|||||
1769 TAACAGCAAGAGCTCAGATTATGTATGAGAGATCCAAAGCAATGCCATG 1818
587 ProValIlePheGlyLysSerSerCysSerGluPheThrArgGluAlaPhe 603
|||||.....:|||||
1819 CCGGTAACTTTGGTAAATCTAGCTGTTCAGAAATTTTCAAGGAGGCTTA 1868
603 eThrProValValTyrHisAsnLysSerProGluPheTyrGluLubhel 620
|||||.....:|||||
1869 TACAGCGCTAGATATATCATACAGCTCTCGATTTTCATGAGCAATCA 1918
620 ySLeuHisLeuProAlaCysValThrGluAsnHisLeuPheThr 636
|||||.....:|||||
1919 AGGTAAACCTTCCTGCTACTTTAACTGACCAATCACTTGTCTTTACT 1968
637 PheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGluThrPr 653
|||||.....:|||||
1969 TTTTATCTATGTTAGTGTCAACAACAAATATCTCTCTTGAACACC 2018
653 oValGlyPheThrTrrPleProLeuLeuGlnHisGlyArgLeuArgThrG 670
|||||.....:|||||
2019 AGTTGATATACATGATACCAATGCTTCAGAAATGAGAGGTTGAAGACTG 2068
670 lYrProPheCysLeuProValSerValAspGlnProProPheSerTyrSer 686
|||||.....:|||||
2069 GCCACTTTTGGCTGCCAGCTTCATTTGAAAAACACACACAGGCTTATCT 2118
687 ValLeuThrProAspValAlaLeuProGlyMetArgTrrValAspGlyN 703
|||||.....:|||||
2119 GTACTGTCTCTCGAGGTTCTCTACTGCTGCAATGGTAGATCAATCA 2168
703 sLysGlyValPheSerValGluLeuThrAlaValSerSerValHisProG 720
|||||.....:|||||
2169 CAAGAGCTGTTTAAATGTTGAAGTGTGCTGTCTGCTATCATACATACAC 2218
720 lAspProTyrLeuAspLysPhePheThrLeuValHisValLeuGluGlu 736
|||||.....:|||||
2219 AAGATCTTATCTTGACAAATTTTGTGCTGTGCTCAATGCTCTGGATGAA 2268
737 GlyAlaPheProPheArgLeuLysAspThrValLeuSerGluGlyAsnVa 753
|||||.....:|||||
2269 CACTGTCTCCAGCTCCGAATTTGGGAGCATGCCAAATCATAGAAAAATACT 2318
753 lGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProG 770
|||||.....:|||||
2319 AAAAAATGATGGAAGACAGATTTTCAGACTGATATCATATCCACGCTGG 2368
770 lProLeuValAlaPheSerHisValLeuAspLysLeuValArgLeu 786
|||||.....:|||||
2369 AACCAAGTGTCCGATTTCTTCATCTTCTGTAGATAAACGTATCTTTTA 2418
787 ValIleArgProProIleLeuSerGlyGlnIleValAsnLeuGlyArgG 803
|||||.....:|||||
2419 GTTATTAGACCTCTGCTCATCTGCGCAAAATGATTAAGCTTAGGCAAGC 2468
803 ValAPheGluAlaMetAlaHisValSerLeuValHisArgSerLeuG 820
|||||.....:|||||
2469 ATCTTTTGAGCGATGCGATCATATTATTAATGATGACTCACAAAACCTGG 2518
820 lAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAlaTyrVal 836
|||||.....:|||||
2519 AAGCAATCATGACCCAGCATGGCAGAAACAGCTTGTGATCATATATAT 2568
837 HisTyrAlaPheArgLeuProGlyThrGluProSerLeuProAspGlyAl 853
|||||.....:|||||
2569 CATATATGTTTCCGCTTACCAAAATCTTACCCTATATCATCATCCACAG 2618
853 aPro.....:ProValThrValGlnAlaAlaThrLeuAlaArgGlys 867
|||||.....:|||||
2619 TCCTGGGGGTGGGAGATCAGTGCATATATGCCCAATGCGTAAATCTG 2668
867 eGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerLysSerSer 883
|||||.....:|||||
2669 CGGTGAGACTGCAGACCTTAATTAAATCGTTTCGAAAGCTTGATAT 2718
884 SerAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSe 900
|||||.....:|||||
2719 AGCAATCCAGATATATCTGAGCTCCAGCTCCAGATGATGAGATTGCG 2768
900 rArgIleLeuAlaSer.....: 905
|||||.....:|||||
2769 ATCAATCATCGGAGTAAAGGTTTAAAGTCAATTCCTGCTGTTAAACA 2818
905 .....: 905
2819 CTGCTGTCCAAAGCTGCCCATGGGAGATCCAAOCCGACGTCAGTGCA 2868
905 .....: 905
2869 GAATCAACACAGGCTATGATGCGAAGTTGTAATCGTATGCTTCCGACAC 2918
906 .....: 906
2919 AGAGAGCTCAAGTTTCTTCAAAACATTAACGGGAGCTTACCAACTAAAA 2968
906 ySLeuLeuHisGluGluLeuAlaLeuGlnTrrValValSerSerSerAla 922
|||||.....:|||||
2969 AGCTTTTTCACGAGGAGCTGGCTTTCAGTGGGTTTTCGAGTGGCAC 3018
923 ValArgGluAlaIleLeuGlnHisAlaTrrPhePhePheGlnLeuMetVa 939
|||||.....:|||||
3019 GTTCGGGAATCACGTTTGCACAAACGCTGTTCTTTTGAATTAATGCT 3068
939 lYsSerMetAlaLeuHisLeuLeuLeuGlnArgGlnArgLeuAspThrPro 956
|||||.....:|||||
3069 AAAGACCATGCTGCACCATTTATCTTATATGATTAACCTTGAGGCTCCAA 3118
956 rGlySerLeuArgPheProGlyArgPheLeuAspThrIleThrAlaLeuVal 972
|||||.....:|||||
3119 GGAAGAGCTGTTTCCAGAAAGCTTTCATGATGATGATTCAGAGCTCTTGC 3168
973 GlySerValGlyLeuGluValIleThrArgValHisLysAspValGluLe 989
|||||.....:|||||
3169 AGCAGATTCGTAGTATATAGTTTCAAGATTTTCAAGAGACACAGAAAT 3218

```

989 uAlaGluHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeu 1006

 3219 GGTGGAGACCTCAATCAAGCCTGCATCTTCTCATGATGATGTTGT 3268
 1006 erLeuValAspArgGlyPheValIlePheSerLeuValAlaHisIleTyrLys 1022

 3269 CTGTTATGGACAGAGATTTGTTTATAGCCTTATAAGCTCCGCTATAAA 3318
 1023 GluValAlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuTh 1039

 3319 CAGGTCTCTCAAGCCTTACTACTATTCAGAAATCCAGAGTTCGTGTGTC 3368
 1039 rLeuArgMetGluPheThrArgIleLeuGlySerHisGluHisIleTyrValT 1056

 3369 CTGGAGCGCGAGATTTCTACGAATCTCTGCACTCATGACACTATGTTA 3418
 1056 hrLeuAsnLeuProGlyCysProLeuSerProAlaSerProSerPro 1072

 3419 CATTAACCTTACCCCTGACCTTACTACTACCCACATCTCCATCCACCT 3468
 1073 SerValSerSerThrSerGlnSerGlnSerThrPheSerSerGluAlaPr 1089

 3469 TCTGTTCTCTGCAACATCTCAGAGTCTCGATTTCTACGAATGTCA 3518
 1089 oAspProLysValIlePheSerPheGluLeuSerGlyProPheArgGlnG 1106

 3519 AGACCAAAAGCATGCAAAATATGTTGATATCCGCTTCGCCCAAC 3568
 1106 hHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlu 1122

 3569 AGCATTTATTTGGCAAGCTGTGTATACAGAGCTGGCTGCTCATTTAGAC 3618
 1123 ProGluAlaGluIleAlaPheLeuLeuHisLysLysAlaIleSerAlaVal 1139

 3619 CCGTAGCTCAAGAGCTGTTGATGATGCAATTAAGAAAGTCATCAATATGCT 3668
 1139 hHisSerLeuLeuGlyGlyHisAspThrAspProArgTyrAlaGluAlaT 1156

 3669 ACGCAATTTACTCTCCAGTCCGACTCAGACCCGCGGACTCTGACCCCTC 3718
 1156 hrValLysAlaArgValAlaGluLeuTyrLeuProLeuLeuSerIleAla 1172

 3719 AGATTAAGGCTCGATGGCCATGTTGATCTATCACTGATGTGTTATTC 3768
 1173 ArgAspThrLeuProArgLeuHisAspPheAlaGluGlyProGlyGlnAr 1189

 3769 ATGGAACTGTACTCAGCTGATGATGATTTACAGAAATCAACAAATCAACG 3818
 1189 gSerArgLeuAlaSerMetLeuAspSerAspThrGluGlyGluAspI 1206

 3819 AGGAGAGCCAAATTTATAGCCATGATGATTAAGAAAGTGAGAGC... 3864
 1206 LeuAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyGlyPro 1222

 3865 ..GGAGATGATAGACCAACCCGTTGCCATGCAATCGCAGGAGCAATCG 3912
 1223 Leu.....AlaProGlySerArgAlaSerIleSerGluGlyPr 1235

 3913 GTCCCTCAACTAACAGGCTGCGACTTTCCTCCTCAGCTCA..... 3954
 1235 oProThrAlaSerArgAlaGlyLysAlaLeuSerAlaGluSerSerArgT 1252

 3955ACGAGGGCAKGCACACATACCTTTCACGAGAAATCAGAGTCGA 4000
 1252 hrLeuLeuAlaCysValLeuThrValLeuLysAsnThrGluProAlaLeu 1268

 4001 GCCCTTTGATCTCTACTTTGGCTTCTCAAAATCGACATGAACAGTT 4050
 1269 LeuGlnArgTTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLe 1285

 4051 CTACAGAAAGGTGTTACAGATCTCTCAGCTTCGACACTAAACCGGCTATT 4100

1285 uAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrLysGlyLysLysA 1302

 4101 AGATCTGCTTTATCTCTGTGTCTTCCCTTTCAGATATAAGGAGAAAAAG 4150
 1302 IaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspMetLys 1318

 4151 TGTTCATACGAATGAATACCTTGACCTTTAAGAATCAAAAGACATGAGA 4200
 1319 AlaArgLeuGluGluAlaIleLeuGlyThrIleGlyAlaArgGlnGluMe 1335

 4201 GCAAGCTTGAAGAAATCTTCTTGAGACATAGGTGCGCAGCAAGAAAT 4250
 1335 tValArgArgSerArg.....GluArgSerPro.....P 1345

 4251 GTTACGCCAGACCGAGAGACCTGACAGAGAAACCCATCTGGAAGTGCT 4300
 1345 hGlyAsnProGluAsnValArgTTrpArgLysSerValThrHisTyrLys 1361

 4301 TTGGAGTCAAGAAATTTGAGGTGAGGAAAGATATGACTCAGCTGCGCT 4350
 1362 GlnThrSerAspArgValAspLysThrLysAspGluMetGluHisGluAl 1378

 4351 CAAGACACAGAGAGCTTCACAAATCAAGACAGAGATTTGAACAGAGAGC 4400
 1378 AlaValAlaGluGlyAsnLeuAlaThrGluAlaSerLeuValValLeuAspT 1395

 4401 ACTGATGTGATGGAAACCTGCTACAGAAACCAATCAATCTTTAGTAA 4450
 1395 hrLeuGluIleIleValGlnThrValMetLeuSerGluAlaArgGluSer 1411

 4451 CATTTAGATTTGTTGTTCCAGACCGTTTCTGTACGGAATCCAAAGAGAGC 4500
 1412 ValLeuGlyValAlaValLeuLysValValLeuTyrSerLeuGlySerAlaG 1428

 4501 ATCTGTTGTTGAGTGTCTAAAGGTCTACTACACAGACAGCGCTGTACCA 4550
 1428 nSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValS 1445

 4551 AAGTCAGTTTATCTCAACACACTGTTTCTTACACAGAGAGCTTGATT 4600
 1445 erLysPheProGluLeuLeuPheGluGluAspThrGluLeuCysAlaAsp 1461

 4601 CAAGTTTCTCTGAACCTTATTTGAAGAAAGACAGACAGAGCTGCTGAT 4650
 1462 LeuCysLeuArgLeuLeuArgHisCysGlySerArgIleSerThrIleAr 1478

 4651 TTATGCTCAGGCTTCTCCGACCTGTAGACAGTACAGATCGTACAAATACG 4700
 1478 gThrHisAlaSerAlaSerLeuTyrIleLeuMetAlaArgGlnAsnPheGlu 1495

 4701 GTCACACCCCGAGTCCCTTACTACTACTAATGAGGCAAACTTTGAGA 4750
 1495 LeuGlnHisAsnPheAlaArgValLysMetGlnValThrMetSerLeuSer 1511

 4751 TTGGGAAATTAACCTTGGCCAGGTGTAATATCGAGTACCAATGTCCACTATCC 4800
 1512 SerLeuValGlyThrArgLysAsnPheSerGluGluHisLeuArgArgSe 1528

 4801 TCTTGGTGGGCAATCTCAGAAATTTTAATGAAGAAATCTTAAAGACGTTT 4850
 1528 rLeuLysThrIleLeuThrTyrAlaGluGluAspMetGlyLeuArgAspS 1545

 4851 TCTTAACACTATATTGACATATGCTGAAAGAGATTCGAAATGAGAGGANA 4900
 1545 erThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIle 1561

 4901 CAACATTTCTCTGATCAAGTCCAGATCTGTTTCAATCTCATATGATT 4950
 1562 LeuThrAspThrValLysMetLysGluHisGlnLysAspProGluMetLe 1578

 4951 CTTTCTATCTGTGAAGAAAGAGAACACAGAGAGATCTCGAAGATGTT 5000
 1578 uIleAspLeuMetTyrArgIleAlaArgGlyTyrGlnGlySerProAspL 1595

```

5001 GATTGATCTAAATGTACAGAAATGGCCAAAGGGTTACACAGACCTCTCCAGAA. 5049
1595 eaurgleuthrtrpleuglnasnmelatalaglylshislaagluengly 1611
5050 .CGATTGACCTGGTGCAGAACATGGCAGCAGACGACCTCAGACAGAAAC 5097
1612 AsnHislaGluAlaAaIacIcysmctValHislaAlaAlaLeuValAl 1628
5098 AATCAATGCTGAACCTGCACAGTCTAGTCCACTCAGCAGACCTGTTGTC 5147
1628 agluTyrlLeuAlaLeuLeugluaspGlnarGHisLeuProValIGlyCysV 1645
5148 TGAATATTGAGCATCTGAGAGACCGGAATATCTCTCTGGGAGATGG 5197
1645 aIserepHeGlnaenIIeserSerAsnValleugluGlnSerAlaIeser 1661
5198 TAACATTTTCAGAAATATTTCATCTAATGTTTGAAGAATCTCGGGCTCA 5247
1662 AspAspIleuSerProaspGluGluGlyPheCysSerGlyLysHisph 1678
5248 GATGATGTGGTATCTCCAGATGAAGAGTATCTGCTCTGGAATACTT 5297
1678 eThrGluLeuGlyLeuValGlyLeuLeugluGlnAlaAlaGlyTyrrPhe 1695
5298 TACTGAGTCCAGACTGTGGGATTACTGGAAACAGCAGCTTCTCTCT 5347
1695 hrnetGlyLeuTyrgluAlaValAsnGluValTyrlsasnleuIle 1711
5348 CTATGGCTGCATGTATGAAGCAGTTAATGAAGTTATCAAAAGTCTTATT 5397
1712 ProIleuGluAlaHisArgAspTyrlsLysLeuAlaAlaValHisgl 1728
5398 CTRATTATAGACTTAATCGGATGCAAGAAACTATCCACATTCATCG 5447
1728 yLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerGlyTyrrp 1745
5448 TAAACTTCAGAAAGCATTCAGCAAAATTTGTTCAATCAGACTGCTGGG 5497
1745 luargValIrhieglyThrTyrrPheArgValGlyPheTyrglyAlaHisSph 1761
5498 AGCGGATTTGGCAGCTATTTGCTGTGTTTATGGAACCAACTTC 5547
1762 GlyAspLeuaspGluGlnIupheValTyrlsGluProSerIleThrIly 1778
5548 GCGGATTTGGATGACAAAGAAATTTGTTACAGAGCCCTCAATTAACCA 5597
1778 sLeuAlaGluIleserHisargLeuGluIupheTyrrThGluArgPheg 1795
5598 ACTTGACAGATATCTCACAGATGGAGGATTTTACGGAGAAAGATTTC 5647
1795 lyAspAspValValGluIleIleLysaspSerTyrrProValAspLysSer 1811
5648 GAGAGGATGGTGAAGTATCAAAACACTTAATCTCTGAGACAACTGT 5697
1812 LysLeuaspSerGlnLysAlaTyrlIleGlnIleThrTyrrValGluProTy 1828
5698 AAATTAATCTCAACAAAGCATATATTCAGATTACTATGGAGCCATA 5747
1828 rPheaspThrTyrgluLeuLysaspArgValThTyrrPheaspArgSnt 1845
5748 CTTTGACACATATGAGATGAAGGACAGATCACTATTTCGACAAAAATT 5797
1845 yrglyLeuArgThrPheLeuPheCysThrProPheThrProaspGlyArg 1861
5798 ACAATCTTCGTCATCTGACTGTACACCCCTTACTTTAGATGGCCGT 5847
1862 AlaHisGlyGluLeuProGluGlnHisLysArgLysThrLeuLeuSerTh 1878
5848 GCCCATGGGAACTTCATGAACAATCAAAAGAAAGCAATCTTGACTAC 5897
1878 rAspHisAlaPheProTyrlIeLysThrArgIleArgValAlcysHisArg 1895

```

```

5898 GTCATGCTCTTCTTATATAATAACAGGTCATGTCATCATMAAG 5947
1895 luGluThrValLeuThrProValGluValAlaIleGluAspMetGlnLys 1911
5948 AAGATATCATCTTAAACCAATTAAGTGTCTATTGAGGACATGACAA 5997
1912 LysThrArgGluLeuAlaPheAlaThrGluGlnAspProProAspAlaTy 1928
5998 AAGACACAGAGATTGGCATTTGCCAACAATCAGGATCCCGACGCCAA 6047
1928 sMetLeuGlnMetValleuGlnGlySerValGlyProThrValAsnGln 1945
6048 AATGCTTCAGATGGTACTCCAGGATCTGTAAGCACCACAGTGAATAG 6097
1945 lProLeuGluValAlaGlnValPheLeuAlaGluIleProGluAspPro 1961
6098 GCCCTTTGGAAGTTGCCAGGTTTTTCTGTGAATAATCAGTACGCCA 6147
1962 LysLeuPheArgHisHisAsnLysLeuArgLeuCysPheLysAspPheCy 1978
6148 AAGCTCTTCAGACATCATTAATAACTGCGACTCTGTTTAAAGATTTC 6197
1978 sLysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProA 1995
6198 TAAAGGTTGTAAGATGCTTAAGAAATAAAGACTTAATTGGCCCG 6247
1995 sPGlnLysGluTyrrHisArgGluLeuGluArg. AsnTyrrCysArgLeuAr 2011
6248 TTCAAAAGAGATATCAAAAGGAATTG...GGGAACATCTCTCCCTTAA 6294
2011 gGluAlaLeuGln. ProLeuLeuThrGlnArgLeuProGlnIleuMetAla 2027
6295 AGAGGCCCTTACAGCCCT...AGATCCAGAAAGTCCCTCAGTTATCCAG 6341
2028 ProThr.....ProProGlyLeuArgAsnSerLeuAsnArgAlaSerPh 2042
6342 CCAGATTGCTGTGCTCCCTGCCACAGACATTCCTTCATCGAATGAGCTT 6391
2042 eArgLysAlaAspLeu 2047
6392 TCGCAAAATGATCTTC 6407

seq_name: gb_pat:AX174569

seq_documentation_block:
LOCUS AX174569 7215 bp DNA linear PAT 04-JUL-2001
DEFINITION Sequence 1 from Patent WO0142296.
ACCESSION AX174569
VERSION AX174569.1 GI:14598200
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 7215)
AUTHORS Lu,P., Garman,J.D. and Candia,A.F.
TITLE Clasp-5 transmembrane protein
JOURNAL Patent: WO 0142296-A 1 14-JUN-2001;
Arbor Vita Corporation (US)
FEATURES
location/Qualifiers
1..7215
/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="full length human CLASP-5 cDNA"
112..6159
/note="human CLASP-5"
/codon_start=1
/protein_id="CAC43733.1"
/db_xref="GI:14598201"
/translation="MTLNSLIDVQLAQLGDFITDDLDVFTFKKERTILOPSLPEGY
ELDPHVDVQVQYIREWLIVNRKNQSGPEICGFKKSGSRKDFHKLTKPFTFSETLGC
SEPAAGPRHLNVLCVSGKGPVTAQDFRLSLQPKRLNLDQVSADEFEKNEE
ARRINRQAEIFALYPSVDEEDVAIVIRVPCREHIGNRILVLTILKFEIETEPFA

```

```
STATDYKERRKISFNHCHDSQDFKGLBAHNPBSVAASQASNAVPSVTYPSDDIT
LVVTEKVLQGDIDDCAPPTVIVIESDGGSKETLEKLOASFCORLQKMPRA
WAPISLSEFFNVSTLEREVTVDSVSGSPVGERRTLAQSRLESLSEENGVPA
FRTSTLSVSEFFKQEDRLSDEDLFKPLADKRSSLODRKYSIPGLRLITSTAPEI
INCCLTPEMLPVKPEPENTRPRKHLEPPTREYVPTVYRNLLYVPOLYNPKL
ASARITTKIOFMGEGEDASNAAPVIFGSGSEPLRKYVITVTHNKSSPDEYEVKIK
LEAKTVHHLHLLFTYHISCOOKOGASVETLLGTSMLPILINERLOTGSTCLPALEK
LEPNTSMHSAEKEVLPONPPIKWABSHKGFVNI EVOAVSSVHTPOHLEKFTTCLHSE
SCVPTPIVLDOKISEMALHEHLEKLISILNNSRLPELHLHLVDLKFQLSVQPMY
IAGQTA NFSOFAFESVVAIANSILNSKDSLSDOGRNCLLASVYHVRLEPVOBDV
KSGAPTALIDPRSYHTGRTSAAAVSSKLOARWSSSNPLGASHADEVKNINS
SKIADRNCSBMSYVCSGSSDAPSSAPARPSAKHHEELALOWVSTGAWKSMACHY
HNMKRDSEPRRPRSDREMDITTYIVNVYSEIALALYKPKENBQAEKNITSLAFEL
YDLUSLMDRGVFNILIRHGCOLSAKLSNLPULLSMRELEFRLICSHHYUNLUFEM
NADAPTPSPCPSISSONSSCSFQDOKIASMFULTSEYRQOHLTGILFTELAAALD
AEDEGISVORKAVASAIHLSLSDHLDRCVKEVYKVAIALYLPVIGIIDLALPOLC
DETVDATRRYRTSGDEDEOGAGAINONVALAIGNNNELTSGVLSVSLPYKQYML
NADTRNLICETIMTKNADOSILIRKNTADLPSTQLNLIIDLFLVCVCEPKKQSS
DKSVTOVLOKSDVKARLEFALLREGARGEEMRRRARNRFGINLENLEMKKQTH
MFOANEKIDTKRAELDOPALISGNLATEBAHIIIDMOENITIOASALDCKDSLGGVL
RYVNSLNCDSSTYTLTHCFATLRALIAKFEDDLFESEVEQCFDLCQVLRHSCSSMD
VTRSOACATLYLLMFSGFATSNFARVMOYTMASLAVGRAPNEBHRLRSRLTIL
AYSEBDTAMOMTPEPTOVELLCLNLSILYDTVMKREOEDEPMLMDIMYIASYOA
SEDLRLTWLQNMMAEKHKKCYTEAAMCLVAAALVAYEYLSLMDHSHYLPVGSYFON
IESNVLESVYSEDTLSPDEGVACAGVETESGLVIGLELOAELFESGLEYETNEY
KLIVPILBAHREPRKLTJTHSKLOARFDSIVNKHKRFEGTYFFVPGFSKFEGLDEO
EYVKEPRITKLEIPELSHLELAPYGCQFGEAEVEYIKDSTPYDKTKLDKNKAYI0ITTY
EYFEYEMKRDVYEFENFNLRRPMYTPPTLELGRVLEABEDYRRNTVLTTHNAP
YIKETISVIOKEEVEVLPRIEVAIEDMKKTKLOLVAIIOEPRDKMLQMVLOQSGVAT
VNOGKLEVAQVLAIEIPADPKLYRHNNKRLCFEKEITRCEAEVKNKRLITTAOR*
OCEIKKNYKLNKENRPIERKPIEYKPIRVSOKRDSFRHSRFRCEYOLDSQS"
```

alignment_scores: Quality: 5258.50 Length: 2075
 Ratio: 3.299 Gaps: 24
Percent Similarity: 76.819 Percent Identity: 51.614

alignment_block:
US-09-736-968a-2 x AX174569 ..

Align seg 1/1 to: AX174569 from: 1 to: 7215

```
36 SetArgArgCysSerSerSerLeuGluValPro...LeuThrGluValVal 51
   :::::::::: :::::::::: :::::::::: :::::::::: ::::::::::
34 ACGGGCGGTGTAACCTTGTAATCTTTCAAGACCTCAGTTTATGACCTGT 83
   ||||||| :::::::::: :::::::::: :::::::::: ::::::::::
51 LgluProLeuAspPheGluAspValLeuLeuSerArgProProAspAlaG 68
   ||||||| :::::::::: :::::::::: :::::::::: ::::::::::
84 GGAGCCAGTGGACTTTGAAGACTCTTGATGACACACCTGACACAGCCTGG 133
   ||:::::::: :::::::::: :::::::::: ::::::::::
68 LuProGluProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
   :::::::::: :::::::::: :::::::::: ::::::::::
134 ATGTCAGACTTGCCCAAGAGCTCGGGGACTTCACTGATGACGACTTGAC 183
   :::::::::: :::::::::: :::::::::: ::::::::::
85 LeuLeuLeuGluInProArgGluCysArgThrThrGluProGlyIlePro 101
   :::::::::: :::::::::: :::::::::: ::::::::::
184 GTGGTGTTCACGCCAAAGAAATAGAGACTTTCAGCCCTCTTGGCCGGA 233
   :::::::::: :::::::::: :::::::::: ::::::::::
101 sAsp...GluIstLeuAspAlaGlnValArgAlaAlaValGluMetTyr 117
   :::::::::: :::::::::: :::::::::: ::::::::::
234 GGAAGGGGTGAACCTGACCCCTATGTCAGGAGCTGTGTTCAAGACCTAC 283
   :::::::::: :::::::::: :::::::::: ::::::::::
117 LeuLAspTrpValIleValHisArgArgTyrGlnTyrLeuSerAlaAla 133
   :::::::::: :::::::::: :::::::::: ::::::::::
284 TCCGTGAGTGGCTATGTCGACCGGAAACCAAGGAAGTCCAGAAATC 333
   :::::::::: :::::::::: :::::::::: ::::::::::
134 TyrSerProValThrThrAspThrGlnArgGluArgGlnIstGlyLeuP 150
   :::::::::: :::::::::: :::::::::: ::::::::::
334 TGTGGCTTTAAAAAGACTGATCTCGAAAGAAATTTTCACAAAGAGCTTCC 383
```

```
150 oArgGlnValPheGluGlnAsp.....AlaSerGlyAspGlu 163
   :::::::::: :::::::::: :::::::::: ::::::::::
384 GAAACAGAGAGTTTGAGTGGAAACCTTGAGTGCATGACACCCGCTGCTC 433
   :::::::::: :::::::::: :::::::::: ::::::::::
163 rSergGlyProGluAspSerAsnAspSerArgArgGlySergGlySero 179
   :::::::::: :::::::::: :::::::::: ::::::::::
434 AGCGAGGCCCCCGCACTTAACGTGCTGTGCGAGCTGTCTGGGAAAGGC 483
   :::::::::: :::::::::: :::::::::: ::::::::::
180 GluAspThrProArgSerSergGlyAlaSerSeriLePheAspLeuArgAs 196
   :::::::::: :::::::::: :::::::::: ::::::::::
484 CCCGTCAC.....GCGTGTGACTTTGACCTCGGAG 515
   :::::::::: :::::::::: :::::::::: ::::::::::
196 mLeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaProG 213
   :::::::::: :::::::::: :::::::::: ::::::::::
516 CCGTCAAGCCTGCAAGCGGCTGAAAAACCTCGCAAGAGAGTGGCGG 565
   :::::::::: :::::::::: :::::::::: ::::::::::
213 LuAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgProPro 229
   :::::::::: :::::::::: :::::::::: ::::::::::
566 AGGACTTTGAGAACGAGAACGAGAGCCCGGAGGACCAAGCGCAGGCC 615
   :::::::::: :::::::::: :::::::::: ::::::::::
230 AlAlaLeuThrLeuTyrProAlaProAspGluAspGluAlaValGluAr 246
   :::::::::: :::::::::: :::::::::: ::::::::::
616 GAGCTCTTGGCCCTTACCACATGAGTGAAGAGAGATGCTGTGAAT 665
   :::::::::: :::::::::: :::::::::: ::::::::::
246 gCysSerArgProGluProProArgGluHisPheGlyGlnArgTyleuV 263
   :::::::::: :::::::::: :::::::::: ::::::::::
666 ACCTCCAGTACCAAGATGTCCCAAGAACCTGGGCAACGAAATATGG 715
   :::::::::: :::::::::: :::::::::: ::::::::::
263 alLysCysLeuSerLeuAspPheGluIleGluIleGluProIlePheGly 279
   :::::::::: :::::::::: :::::::::: ::::::::::
716 TCAAGTGTGTAACCTTGAAGTTCGAGATTGAATGAAACCCCTGTTTCC 765
   :::::::::: :::::::::: :::::::::: ::::::::::
280 lLeuAlaAlaLeuTyrAspValArgGluIstLysLysIleSergLysAsn 296
   :::::::::: :::::::::: :::::::::: ::::::::::
766 AGCATTTGCCCTTACGATGTTAAAGAAAGAAAGATCTCAGAAATATT 815
   :::::::::: :::::::::: :::::::::: ::::::::::
296 eTyrThrAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAlaHisG 313
   :::::::::: :::::::::: :::::::::: ::::::::::
816 TCACGTGTACCTGAACCTGTGACCAAGATTTGCGAGCTGCACA 865
   :::::::::: :::::::::: :::::::::: ::::::::::
313 lYThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSeryAl 329
   :::::::::: :::::::::: :::::::::: ::::::::::
866 GCGCTTCAGTGGCGGCATCAAGTCAAGGAGATCGAGTGTCTGCANCT 915
   :::::::::: :::::::::: :::::::::: ::::::::::
330 ThrTyrProSerProAspIlePheLeuValIleLysLeuGluLysValLe 346
   :::::::::: :::::::::: :::::::::: ::::::::::
916 ACCTACCCGCTCCTCAGACATCTACCTGGTAGCAAGATTGAAAAAGTCT 965
   :::::::::: :::::::::: :::::::::: ::::::::::
346 uGlnGlnGlyAspIleSergGluCysGluProTyrMetValLeuLysG 363
   :::::::::: :::::::::: :::::::::: ::::::::::
966 GCAGCAGGAGATATTGGAGACTGTGCAGAGCCCTTACAGGTTATCAAG 1015
   :::::::::: :::::::::: :::::::::: ::::::::::
363 LuValAspThrAlaLysAsnLysGluLysLeuGluLysLeuArgLeuAla 379
   :::::::::: :::::::::: :::::::::: ::::::::::
1016 AAAGTATGATGGTGAAGAGTAAAGAAAGATTGAAGAAACTAAACTCCA 1065
   :::::::::: :::::::::: :::::::::: ::::::::::
380 AlaGluGlnPheCysThrArgLeuGluArgTyrArgMetProPheAlaTr 396
   :::::::::: :::::::::: :::::::::: ::::::::::
1066 GCTGATCTCTTCCACACGCTTTGGGGAATACCGAGAGCCCTTGGCTG 1115
   :::::::::: :::::::::: :::::::::: ::::::::::
396 pThrAlaValHisLeuAlaAsnIle.....ValSerSerAlaGlyInL 411
   :::::::::: :::::::::: :::::::::: ::::::::::
1116 GGCACCCATTAAGCTTATCAAGCTTCTTCAATGTCTCCACCTTGAGAG 1165
   :::::::::: :::::::::: :::::::::: ::::::::::
411 euAspArgAspSerAspSer.....GluGlyGluArg 421
   :::::::::: :::::::::: :::::::::: ::::::::::
1166 AGCTAAGCATGATGTGACCTGTGCTGGGAGAGAGCCAGTGGGTGAACG 1215
   :::::::::: :::::::::: :::::::::: ::::::::::
422 ArgProAlaIleTrpThrAspArgArgArgArgGlyProGlnAspArgAla 438
   :::::::::: :::::::::: :::::::::: ::::::::::
1216 AGGACATTGGCCCAATCTAAGAGGCTT.....TGTGAAGAGGCTT 1256
   :::::::::: :::::::::: :::::::::: ::::::::::
438 rSergGlyAspAlaCysSerPheSergGlyPheArgProAlaThrLeu 455
```



```
111 ..... |||:||||: ::|||:
1257 CTCCTGAGGAAATGGGTTGGATCCAACTCAAAACCTCCACTGTA 1306
455 hrValThrAsnPhePheIysGlnGluAlaGluSerAspGluAsp 471
   ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
1307 GCGTTGAGCTTTTTCAGCGAGGAGAGATGCCCTTAGCGAGGAGAC 1356
472 LeuPheIysPheLeuAlaAspMetArgArgProSerSerLeuLeuArg 488
   |||:|||||:|||||:|||||:|||||:|||||:|||||:
1357 TTATTCAAGCTTTTAGCTGACTACAAAAGATCATCTTACAGAGACG 1406
488 GLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPro 505
   ::|||:||||:||||:||||:||||:||||:||||:
1407 AGTCAGAGCAATTCAGGCTTGCTAGACGAGGAGATTCTACAGCTCAG 1456
505 LuasnProHisPheCysLeuSerProGluLeuLeuHisIleLysProTyr 521
   |||:||||:||||:||||:||||:||||:||||:
1457 AGATCATCAATTCGTGCTGACTCTGTGAATGCTGCCCGTGAACCCCTTT 1506
522 ProAspProArgGlyArgProThrLysGluIleLeuGluPheProAlaAr 538
   |||:||||:||||:||||:||||:||||:||||:
1507 CCTGAAAACGGACACGCCGCCACAAAGAGATTGTGGAATTTCCACACG 1556
538 GGUValThrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTyrP 555
   |||:||||:||||:||||:||||:||||:||||:
1557 AGAAGTATATGTCCTCCACACTGTGTACAGAAACCTCTCTATGTCACC 1606
555 rOHISerLeuAsnPheSerSerArgGlnGlySerValArgAsnLeuAla 571
   |||:||||:||||:||||:||||:||||:||||:
1607 CACGAGGCTCACTTGTGAACAACTAGCATACGCCGGAACATTACA 1656
572 ValArgValGlnTyrMetThrGlyGluAspProSerGlnAlaLeuProVa 588
   ::|||:||||:||||:||||:||||:||||:||||:
1657 ATAAAGATCCAGTTTATGTGTGAGAGAGATGCTAGCAATGGAGTCCGGT 1706
588 ILePheGlyLysSerSerCysSerGluPheThrArgGluAlaPheThrP 605
   |||:||||:||||:||||:||||:||||:||||:
1707 CATCTTTGGAAATCCAGCGGCGCTGAATTTCTGCAGGAAGTGTACACAG 1756
605 roValValTyrHisAsnLysSerProGluPheTyrGlnGluPheLysLeu 621
   |||:||||:||||:||||:||||:||||:||||:
1757 CTGTTACTTACCATTAATAGTCTCTGACTTTTATGAAGAAGTGAATAAT 1806
622 HisLeuProAlaCysValThrGluAsnHisLeuLeuPheThrPheTy 638
   |||:||||:||||:||||:||||:||||:||||:
1807 AAGCTCCCGGCTAGCTACAGTAATCACACCTCCTGTTCACTTCTA 1856
638 rHisValSerCysGlnProArgProGlyThrAlaLeuGluThrProValG 655
   |||:||||:||||:||||:||||:||||:||||:
1857 CCATATACAGCTGTACGAGAGCAAGAGAGCCCTCGTGAACCTCTCTGG 1906
655 LysPheThrTrpIleProLeuLeuGlnHisGlyArgLeuArgThrGlyPro 671
   |||:||||:||||:||||:||||:||||:||||:
1907 GATATTATGCTGCTCCCAATTTCTTTAAATGACSTCTTCAACTGAGTCC 1956
672 PheCysLeuProValSerValAspGlnProProSerSerTyrSerValLe 688
   ::|||:||||:||||:||||:||||:||||:||||:
1957 TACTGTCTCCAGTTCCTTGGAATAATTTGCCACCCCAACTACCTCATGCA 2006
688 uThrProAsp.....ValAlaLeuProGlyMetArgTrpValAspG 702
   ::|||:||||:||||:||||:||||:||||:||||:
2007 TTTCTGTGAGAGAACTCCCATTTACAGAAATCCCTCCATTAAAGTGGGCTAG 2056
702 LysHisLysGlyValPheSerValGluLeuThrAlaValSerSerValHis 718
   |||:||||:||||:||||:||||:||||:||||:
2057 GACTAAGAGGAGTATTATTAATTAAGTGCAGAGCTGTTTCTCTGTATCAC 2106
719 ProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValLeuG 735
   |||:||||:||||:||||:||||:||||:||||:
2107 ACCCAGGACAAACACACTGAGAAAGTTCTTCACCCTTGCCACCTCCCGGA 2156
735 u...GluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGluG 751
   |||:||||:||||:||||:||||:||||:||||:
2157 GAGCCAGGTGACTTCCCATCCGCGTGTGATCAGAAATACACGAGA 2206
751 LysAsnValGlnGlnGluLeuArgAlaSerLeuAlaLeuArgLeuAla 767
   ::|||:||||:||||:||||:||||:||||:||||:
2207 TGGCGCTGAGCATGAGCTGAAAGCTCAGCATCATCTGCTGAACCTCC 2256
768 SerProGluProLeuValAlaPheSerHisValLeuAspLysLeuVal 784
   |||:||||:||||:||||:||||:||||:||||:
2257 GCGCTGAGCCGCGTGTCTTCCGTGACACCTGAGTGTGAGACACCTCTT 2306
784 LArgLeuValIleArgProProIleIleSerGlyGlnIleValIasnLeuG 801
   ::|||:||||:||||:||||:||||:||||:||||:
2307 CCAGCTCCGCGACCCCATGCTGCTGCGTCGCGACAGACCAACTCTT 2356
801 LArgGlyValaPheGlnAlaMetAlaHisValIleValSerLeuValHisArg 817
   |||:||||:||||:||||:||||:||||:||||:
2357 CCCAGTTTCCCTTCGAGTCCGTGTGGCCATGCCACACAGTGTGCACAC 2406
818 SerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAl 834
   |||:||||:||||:||||:||||:||||:||||:
2407 AGCAAGAGACTGAGCAGAGCAGCATGGAGAGACTGCGCTGCTGCTTC 2456
834 aTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPro. 850
   ::|||:||||:||||:||||:||||:||||:||||:
2457 CTACGTCACCTACGTCTTCCGCTGCAGAGGTGCAGAAAGGATGTGCCCA 2506
851 ..AspGlyAlaProProValThrValGlnAlaAla.....ThrLeu 863
   |||:||||:||||:||||:||||:||||:||||:
2507 AGTCAGGCTCCACCTGCCCTCTAGACCTTCGGAGCTACACACAGTAT 2556
864 AlaArgGlySerGlyLysArgProAlaSerLeuTyrLeuAlaArgSerLysSe 880
   ::|||:||||:||||:||||:||||:||||:||||:
2557 GCGCGCATCATCAGCTGCTGCTGAGTTCAAAGCTGTGCACAGCCCGGCT 2606
880 rIleSerSerSerAsnProAspLeuAlaValAlaProGlySerValAspA 897
   ::|||:||||:||||:||||:||||:||||:||||:
2607 GATGACACACAGTAACCCAGACCTCGCGGGGACACACTCCGACGACAGC 2656
897 spGluValSerArgIleLeuAlaSerLysLeu..... 907
   ::|||:||||:||||:||||:||||:||||:||||:
2657 AGGAATGTGAAGAAACATCATGCTTTCAAAGATGCCGATGCCAATGCAGC 2706
907 ..... 907
2707 CGAATGCTTACTATTGCTGTGACAGTAGTAGTCCAAAGTTCACTGC 2756
908 ..... LeuHisGlnGluLeuAlaLeuG 915
2757 AGCCCCAAGGCCAGCCGCAAAAGCATTTCCATGAGGAGCTTGCCCTTC 2806
915 lntPrValIalSerSerSerAlaValaArgGluAlaIleLeuGlnHisAla 931
   |||:||||:||||:||||:||||:||||:||||:
2807 ACATGCTGTGTCAGCACCGGA..... 2826
932 TrpPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuLe 948
   |||:||||:||||:||||:||||:||||:||||:
2827 ..... ATGCTGAAGAGCATGCGCCAGCAGCTACATAA 2858
948 uGlyGlnArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLe 965
   ::|||:||||:||||:||||:||||:||||:||||:
2859 CATGGACAACGCGAGAGCTTTTCGGAGAGCTGTTTTTCTGACCCGTTTCA 2908
965 euAspAspIleThrAlaLeuValaGlySerValaGlyLeuGluVala..... 979
   ::|||:||||:||||:||||:||||:||||:||||:
2909 TGGATACACTAATCTACTATTGTTAATGTGGTACCTCCGGAATTTGCACCC 2958
980 ...IleThrArgValHisLysAspValaGluLeuAlaGlnHisLeuAsnAl 995
   ::|||:||||:||||:||||:||||:||||:||||:
2959 CTTTATGTAACACACAGAAAGAAATGAAACAGCGGAAAGATGAAACAT 3008
995 aSerLeuAlaPhePheLeuSerAspLeuLeuSerLeuValaAspArgGlyP 1012
   |||:||||:||||:||||:||||:||||:||||:
3009 CAGCGTGGCTTCTCTGTATGACCTTCTCTCCCTCATGATCGGGCT 3058
```

```

1012 heValPheSerLeuValArgAlaHisTyrLysGlnValAlaIathrArgLeu 1028
|||||.....| | | | | : : : | | | | | : : : | | |
3059 TGGGTTAACTCATGACATATTATGACGCCAGCTGTCAGCCAAGCTC 3108
1029 GlnSerProAsnProAlaIleLeuThrLeuArgMetGluPheTh 1045
: : : : : | | | | | : : : : : | | | | | : : : : :
3109 AGTAACTTCCA.....ACGCTCATTTCCATGAGGCTTAAGTTCCCT 3149
1045 rArgIleLeuCySerHisGlnHisTyrValThrLeuAsnLeuProCysC 1062
|||||.....| | | | | : : : : : | | | | | : : : : :
3150 GAGAAATCCTCTGATGCCATGACATTAACCTCAATCGAACTTTTTTTA 3199
1062 ySProLeu...SerProAlaSerProSerProSerValSerSerThr 1077
: : : : : | | | | | : : : : : | | | | | : : : : :
3200 TGAATGCTGATACGCTCCACACATCTCTGCTTCCTCCATATCT..... 3243
1078 ThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLysValTh 1094
|||||.....| | | | | : : : : : | | | | | : : : : :
3244 ...TCCGAGAAGTCAAGCTCCTGCTCCAGCTTCAGAGACCAGAAATCGC 3290
1094 rSerMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaG 1111
|||||.....| | | | | : : : : : | | | | | : : : : :
3291 CAGCATGTGTGATCTGACTTCCGAGTACCCGACAGCACTTCTCAGCG 3340
1111 LysLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGly 1127
|||||.....| | | | | : : : : : | | | | | : : : : :
3341 GGCTCCTTTCACAGAACTGGCTGCTGCTGCTGATGCCGAGAGGGAAGGA 3390
1128 AlaPheLeuHisTyrLysAlaIleSerAlaValHisSerLeuLeuCy 1144
: : : : : | | | | | : : : : : | | | | | : : : : :
3391 ATCAGCAAACTACAAAGAAAGCTGTACAGTCAATTCAACAGCTGCTAAG 3440
1144 sGlyHisAspThrAspProArgGlyAlaGluAlaThrValLysAlaArgY 1161
|||||.....| | | | | : : : : : | | | | | : : : : :
3441 TTCTCAGCAGCTGAGCCACGCTGTGTCAAAACGAGGTGAAGTCAAAA 3490
1161 alaIaGluLeuTyrLeuProLeuLeuSerIleAlaIaArgAspThrLeuPro 1177
|||||.....| | | | | : : : : : | | | | | : : : : :
3491 TCGCGCGCCTTTACCTACCTTTAGTTGGATCACTTTGGATGCTTTGCCA 3540
1178 ArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaSe 1194
: : : : : | | | | | : : : : : | | | | | : : : : :
3541 CACCTCTGTACTTACAGTTGCAGATACTCGCAATACCGCACACAGTGG 3590
1194 rMetLeuAspSerAspThrGluGlyGlnGlyAspIleAlaGlyThrIleA 1211
: : : : : | | | | | : : : : : | | | | | : : : : :
3591 C.....TCGGATGAAAGAACAAAGAGA.....GCCGGTGCATTAA 3625
1211 snProSerValAlaMetAlaIleAlaGlyLysProLeuAlaProGlySer 1227
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
3626 ACCAGAAATGCGCTGCGCATGACGAGGAATATTCAATTGAAAACA 3675
1228 ArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlyCysAl 1244
: : : : : | | | | | : : : : : | | | | | : : : : :
3676 AGTGAATAGTGTCTTCCTCCCTGCCC.....TATTAAGACATCAACAAT 3719
1244 alaSerAlaGluSerSerArgThrLeuLeuAlaCysValLeuTyrValL 1261
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
3720 GGTGAACGGCGACATCTGCAACCTCATGATCTCTCTCTGGATCA 3769
1261 euLysAsnThrGluProAlaLeuLeuGlnArgThrAlaThrAspLeuThr 1277
: : : : : | | | | | : : : : : | | | | | : : : : :
3770 TGAATAATGCTGATCAGAGCCTCATTAAGAAAGTGAATTCGTGCTGCA 3819
1278 LeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaI 1294
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
3820 TCAACGACAGCTCAACAGATTTTAGATCTACTTTTCATCTGTGTATTAG 3869
1294 aPheGluTyrLysGlyLysAlaPheGluArgIleAsnSerLeuThrP 1311
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
3870 TTTTGTAGTAAAGG3AAACAGAGTTCTGACAAAGTCACTACCCCAAGTCC 3919

```

```

1311 heLysLysSerLeuAspMetLysAlaArgLeuGluAlaIleLeuGly 1327
: : : : : | | | | | : : : : : | | | | | : : : : :
3920 TGCAGAAAGTCAAGGATGTCAAGGCCGCTGAGAGAGCTTGTGCTGGT 3969
1328 ThrIleGlyAlaArgGlnGluMetValArgArg.....SerArgG1 1341
| | | | | | | | | | | | | | | | | | | | | | | | | | |
3970 GGGGAAGGGGGCCAGAGGGAGATGATGGCCGCCCGCGCTCCAGGAAACA 4019
1341 uArgSerProPheGlyAsnProGluAsnValArgTrpArgLysSerValT 1358
| | | | | | | | | | | | | | | | | | | | | | | | | | |
4020 CCGATTTCGA...GGCCTAAATGAATAATTGAGATGAGAAAGACAGAGA 4066
1358 hrHisTrpLysGlnThrSerAspArgValAspLysThrLysAspGluMet 1374
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4067 CACATTGGCGGCAAGCTAATGAGAGCTAGATTAATAACAAAGCCGAGTTA 4116
1375 GlnHisGluAlaLeuValGluLysAsnLeuAlaThrGluAlaSerLeuVa 1391
: : : : : | | | | | : : : : : | | | | | : : : : :
4117 GATCAAGAAAGCCTTATCATGCGCAATCTGGCTACAGAAACACATTTAAT 4166
1391 ValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGluA 1408
: : : : : | | | | | : : : : : | | | | | : : : : :
4167 CATCCGTGATATGACAGAAACATATATCCAGGCGAGTCTGCGTCTGACT 4216
1408 IaArgGluSerValLeuGlyAlaValLeuLysValValLeuTyrSerLeu 1424
: : : : : | | | | | : : : : : | | | | | : : : : :
4217 GTAAAGACAGCTGCTGGAGGTGTCTGAGGGGTCTGTGAATTTCTGTG 4266
1425 GlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnAr 1441
: : : : : | | | | | : : : : : | | | | | : : : : :
4267 AACGTGATACAGATTAACACCACTGACTCATCTGCTTTGCAACACTCGG 4316
1441 gAlaLeuValSerLysPheProGlnLeuLeuPheGluGluAspThrGluL 1458
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4317 TCTCTCATCGCAAGTTTGGAGACTTCTCTGCAAGAGAGGGTGAAC 4366
1458 euCysAlaAspLeuCysLeuArgLeuArgHisCysGlySerArgIle 1474
| | | | | | | | | | | : : : : : | | | | | : : : : :
4367 AGTGTTCGACCTATGATGACCAAGTCTGCAACACTGCAAGCAGCATG 4416
1475 SerThrIleArgThrHisAlaSerAlaSerLeuTyrLeuMetArgG1 1491
: : : : : | | | | | : : : : : | | | | | : : : : :
4417 GATGTACCCGAGGCAAGCCTGTGCACCTTTACCTCTCATGAGGTT 4466
1491 nasPheGluIleGlyHisAsnPheAlaArgValLysMetGlnValThrM 1508
: : : : : | | | | | : : : : : | | | | | : : : : :
4467 CAGTTTGGAGCCACCACTAATTTTGCAGAGTAAAGATGCAAGTAACA 4516
1508 etSerLeuSerSerLeuValGlyThrGlnAsnPheSerGluGlnHis 1524
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4517 TGTCCCTGGCATCTTTGGTGGGAAGACACCAAGCTTTAATGAAGAAGC 4566
1525 LeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluGlnAspMetG1 1541
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4567 CTGAAAGATCTCTTGAGACAATTTTGGCCTATTTCAGAAAGACACAGC 4616
1541 YLeuArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnL 1558
: : : : : | | | | | : : : : : | | | | | : : : : :
4617 CATGCAAGATGACTCTCTTCCACCCACAGTGGAGAGAACTTCTGTATATC 4666
1558 euHisMetIleLeuThrAspThrValLysMetLysGlnHisGlnGluAsp 1574
| | | | | | | | | | | | | | | | | | | | | | | | | | |
4667 TGAATAGCATCTTATATGACACAGTGAAGAAAGAGGAAATTTCAAGAAAT 4716
1575 ProGluMetLeuLeuAspLeuMetLysArgIleAlaArgGlyTyrGlnG1 1591
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4717 COTGAGATGCTTATGAGATCATGTACAGAAATTTGCCAAGAAATTCAGAGC 4766
1591 ySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisA 1608
| | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
4767 ATCTCTGATCTGGCGGTGACCTGCTCCAGAAATCATGCAAGAAACACA 4816
1608 IaGluLeuGlyAsnHisAlaGluAlaIleGlnCysMetValHisAlaIa 1624

```

```

4817 CCAAGAGAAGTGTCTACACGAGGCTGCATGTGCTGGTGCAAGCCGCT 4866
1625 AlaLeuValAlaGluThrLeuAlaLeuLeuGluAspGlnArgHisLeuP 1641
4867 GCGTGTAGTGGCTGATATGTGACATGTGAGACCAACAGCTACTGCTGCC 4916
1641 oValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuGluGlnS 1658
4917 CGTGGGAGAGTTCAGCTTCAGAAATTTCTTCATGTGCTGGAGAGAT 4966
1658 eValIleSerAspAspIleLeuSerProAspGluGluGlyPheCysSer 1674
4967 CTGTGGCTCTGAGAGACACCTGTCACTGACAGAGATGGGGTGTCCGA 5016
1675 GlyHisHisPheThrGluLeuGlyLeuValGlyLeuLeuGluGlnAla 1691
5017 GGGCAAGTACTTCACCGAGAGTGGCTGTGAGGCTCTCTGGAGAGGCGC 5066
1691 acGlyThrPheThrMetGlyGlyLeuThrGluAlaValAsnGluValThr 1708
5067 GAGAGCTCTTCAGCAGCGAGGCTTATATGAGACAGTTAATGAGGTCTACA 5116
1708 yAsnLeuIleProIleLeuGluAlaHisArgAspTyrIleCysLeuAla 1724
5117 AGCTGGTCATCCCATCTCAGAAAGCCATCGAAGATTCGGAGAGCTGACA 5166
1725 AlaValHisGlyLeuLeuGlnGluAlaPheThrLysIleMetHisGln 1741
5167 CTCACATCACAGCAAGCTGCAGAGAGCTTCGACAGCATGCTTACACAGA 5216
1741 rSerGlyThrGluArgValPheGlyThrTyrPheArgValGlyPheArg 1758
5217 TCAT.....AAGAAATGTTTGAACCTACTCTCGAGTTGGTCTTGG 5260
1758 LyAlaHisPheGlyAspLeuAspGluGlnGluPheValTyrIleGluPro 1774
5261 GATTCCAAAATTTGGGATTTGGATGAACAGAGATTGTCTACAAAGAGCCT 5310
1775 SerIleThrLysLeuAlaGluIleSerHisArgLeuGluGluPheThr 1791
5311 GCATTTACCAAGCTTCCTGATCTCACAATAGCATGAGGACATTTATG 5360
1791 rGluArgPheGlyAspAspValValGluIleIleLysAspSerTyrPro 1808
5361 TCATATGTTTGTGTCAGAAATTTGTGAGAGATTAAAGATCCACCTCG 5410
1808 aLAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyr 1824
5411 TGGACAAACCAAGTTGGATCTTACACAGGCTTACATACGATCATTCTT 5460
1825 ValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTyrPh 1841
5461 GTGGAGCCCTACTTTGATGTAGTATGAGATGAAGACAGGCTCATACAT 5510
1841 eAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProPheThrP 1858
5511 TGAAGAAATTTCAACCTCCGAGGTTCAATGATCACACCCCTTCACCC 5560
1858 roAspGlyArgAlaHisGlyLeuLeuProGluGlnHisLysArgLysThr 1874
5561 TGGAGGGGCGGCTCGGGAGAGCTGCATGAGCAGTACAGAGGATCGCG 5610
1875 LeuLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArgVa 1891
5611 GTCTGTACACATATGACGCGCTTCCCTACATCAACAGCAGATCGCG 5660
1891 LcysHisArgGluGluThrValLeuThrProValGluValAlaIleGluA 1908
5661 CATCCCAAAAGAGAGATTGTTTGAACCCGATTTGAAGTTGCCATTGAAG 5710
1908 spMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluGlnAspPro 1924
|||||..... ||| .....

```

```

5711 ACATGAGAGAGAACCTCGAGTTAGCAGTTGCCATTTACACGAGCCG 5760
1925 ProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyProTh 1941
5761 CCGTGTCAAAAGATGCTTACATGTGTGCTGCAAGGCTCTGGGAGCTAC 5810
1941 rValAsnGlnGlyProLeuGluValAlaGlnValPheLeuAlaGluIleP 1958
5811 TGTAAATCAGGAGCACCTGGAAGTAGCCCAAGTGTGTTTGGCTGAATTC 5860
1958 roGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeuCysPhe 1974
5861 CTGTGATCCAAACTCTATGACATCACCAACAAGTTGAGGTATGCTTT 5910
1975 LysAspPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLe 1991
5911 AAGGAATTCATCAGATGTGTGTCGAACCTTACAGAAAAACAAGCGCT 5960
1991 uIleGlyProAspGlnLysGluTyrHisArgGluLeuGluArgAsnTyrC 2008
5961 CATCAGGAGCAGACAGAGCAATATCAGACAGAACTCAAAAAACAATATA 6010
2008 yArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGln 2024
6011 ACAACCTAAAGAACCTCAGGCCAATGATCGAGCGAAATTCACAGAA 6060
2025 LeuMetAlaPro.....ThrProGluLeuArgAsnSerLeuAs 2038
6061 CTGTCAAGCCCAATATTCAGAGTTGAGAGTCAAAAGAGGAGCTCTTCCA 6110
2038 nArgAlaSerPheArgLysAlaAsp 2046
6111 CAGATCTAGTTTCAGGAATGTGAA 6135

seq_name: gb_ov:AY016023

seq documentation block:
LOCUS AY016023 167215 bp DNA linear VRT 01-FEB-2002
DEFINITION Spherooides nephelus alpha globin gene cluster, complete sequence.
ACCESSION AY016023
VERSION AY016023.1 GI:18463963
KEYWORDS
SOURCE
ORGANISM
Spherooides nephelus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Spherooides.
REFERENCE
1 (bases 1 to 167215)
Flint,J., Tufarelli,C., Peden,J., Clark,K., Daniels,R.J.,
Hardison,R., Miller,W., Phillips,S., Tan-Un,K.C., McMorrow,T.,
Frampton,J., Alter,B.P., Fritschau,A.M. and Higgs,D.R.
Comparative genome analysis delimits a chromosomal domain and
identifies key regulatory elements in the alpha globin cluster
Hum. Mol. Genet. 10 (4), 371-382 (2001)
JOURNAL
MEDLINE 21096913
PUBMED 11157800
REFERENCE
2 (bases 1 to 167215)
Flint,J., Tufarelli,C., Peden,J., Clark,K., Daniels,R.J.,
Hardison,R., Miller,W., Phillips,S., Tan-Un,K.C., McMorrow,T.,
Frampton,J., Alter,B.P., Fritschau,A.M. and Higgs,D.R.
Direct Submission
Submitted (06-DEC-2000) MRC Molecular Haematology Unit, Institute
of Molecular Medicine, John Radcliffe Hospital, Oxford, Oxon OX3
9DS, UK
FEATURES
source
1..167215
location/Qualifiers
/organism="Spherooides nephelus"
/db_xref="taxon:39110"
1510..1530
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=c

```

```
repeat_region 2315..2350
/note="C-rich; low complexity"
repeat_region 2643..2663
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 3620..3676
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 4231..4284
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 4638..4678
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 5366..5600
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 5840..5861
/note="AT-rich; low complexity"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 6334..8354
/note="AT-rich; low complexity"
repeat_region 9950..9970
/note="AT-rich; low complexity"
repeat_region 10777..10802
/note="AT-rich; low complexity"
join(<10872..11080,11797..>11973)
/gene="unkempt"
/product="unkempt"
/codon_start=1
/protein_id="AAL73045.1"
/db_xref="GI:18463964"
/translation="YLKEFRTQCPFLQHKCTQHPPTCFNMHFLNQRPRRIRRD
GTENVSPDYCTKDETTGCPDGDCCPYLHFTTGTEKKYHLRYKKTGTCTJHEHDAR
GHCVKNGLHCAFAHGPDLRPPYDI"
17726..17763
/note="A-rich; low complexity"
repeat_region 19346..19371
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=c
20068..20121
/note="AT-rich; low complexity"
complement(<21113..>22747)
/gene="IGFALS"
/product="IGFALS"
complement(<21113..>22747)
/gene="IGFALS"
complement(<21113..>22747)
/gene="IGFALS"
/codon_start=1
/product="IGFALS"
/protein_id="AAL73046.1"
/db_xref="GI:18463965"
/translation="ACARGCTCOHDDYSLELANYCSAONLTQVPPDMPSTSLMDV
NLFASLPAAFSKDIHLEFLNLSGOLVTLDPALGSLAHILHERRIRIVLDGAI
FONMPKIASLSTNNOLSRIDRLFAGLSHMTLNGRSIAVILPPTVRODQIGKEI
LIYGNRLAYIQPOLFOFNVYELKELDLSGNLTAKYIKANVYKTLKOKLIYAONQITV
VPRFVGMKSLRMIDLTNNRLTSLHEDEFGVLTLLVLRVSNSTIGIKPRFRDIQY
LEELRLSHNRIRVLEGEVFEELRLELEENNRIOEAVSGFTGLSHAVTNLSGSC
FTLIPDOVFRGLSLHSLHDLRCGLTRVTQAFSGLSLRRLFTLQNNNTSVYEROSFV
DIVGLIGLDSFNKLEVLITGOTSGGLKNLEYLLSNDCRHLFLONGSRLLRLRLID
LRANELTSIVPEAEGLLEKLLSGNRMRCCGALPLRNVSLRNALVVPQVETNAHGE
EPPTIITVNNITCHSPRLTGODLRDVESSVQSC"
```

```
repeat_region 24332..24356
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 24525..24554
/note="GC-rich; low complexity"
repeat_region 24862..24895
/note="AT-rich; low complexity"
repeat_region 26315..26340
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 26915..26951
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=cgg
repeat_region 27240..27272
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 27487..27507
/note="AT-rich; low complexity"
repeat_region 28056..28080
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 28159..28189
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=cag
repeat_region 30217..30260
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 31181..31214
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 33149..33186
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 33186..33241
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=gga
repeat_region 33298..33379
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 33966..34015
/note="T-rich; low complexity"
repeat_region 35457..35500
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 35965..35993
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 37605..37637
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=tcg
repeat_region 37782..37804
/note="simple repeat"
/rpt_type=tandem
/rpt_unit=ca
repeat_region 40208..40383
```

```
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tca
40386..40562
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=tcta
41620..41645
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=ca
41968..41994
/note="simple repeat"
/rpl_type=tandem
/rpl_unit=lg9
44729..44773
/note="simple repeat"
/rpl_type=tandem
```

```
alignment_scores:
  quality: 4734.00      length: 5488
  ratio: 2.649          gaps: 82
  percent similarity: 32.562  percent identity: 25.765
```

alignment block:
US-09-736-968a-2 x AY016023/rev ..

Align seg 1/1 to reverse of: AY016023 from: 1 to: 167215

```
45 ValProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134857 TTGCCCCGTGACAGAGGTGTGGAGCCGGTGGATTGTGACGATTATGTGAG 134808
   61 uSerAlaGProProAspAlaGluProGlyProLeuArgAspLeuValGluP 78
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
134807 CACCCATGCTCTCGTGGGTGGAGCTGTCCCTCAGAGAGCTTGTGAGT 134758
   78 heProAlaAspAspLeuGluLeuLeuGluProArgGluCysArgThr 94
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134757 TCCCCCGACAGACCTGAGACTCTCCACCTGACAAAGAGTGCACACAC 134708
   95 ThrGluProGlyIleProLysAspGluLys..... 104
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134707 CTGGAGGCCCTCGCTGCTGTAAGAGAGATTGTTCCTCTGTGCC 134658
   104 ..... 104
134657 AACCAATGCTCACACACTCTGTCTTTCCTTTTAAACAACACACA 134608
   104 ..... 104
134607 CACACACACACATACACAGATGATGATTACATATTGAGTGGAGG 134558
   104 ..... 104
134557 TGAACATGACAGGACCAAGGCTTTTGAAGTGGCTGAATAACAAC 134508
   104 ..... 104
134507 ACACTTCTGATTTAAAGACATTTCTGCTCAATTATTCATGTGTTGTC 134458
   105 .....LeuAspAlaGluValArgAlaAlaValGluMetTyrIle 117
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134457 TGTGTAGCTGACTGATCCAGAGTGCAGACACCTGTGCTAGTCTACACA 134408
   118 GluAspTyrValIleValHisArg..... 125
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134407 GATGACTGGCTCATTCATTCAAAGAAAGTAAGTCTACATGTGAGGAGA 134358
   125 ..... 125
134357 CCGTAAATGTTCTCTCTGACAGACAGCCCTTTTATCTGTGATTG 134308
```

```
126 .....ArgTyrGlnTyrLeuSer 131
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134307 CTTCGTTATTAACTCGCTCTCTTTTATGCTAGATACCGCACTACAGC 134258
   132 AlaAlaTyrSerProValThrThrAspThrGluArgGluArgGlnLysG 148
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134257 ACCACTTTACAGCCTCAGAGCTCAGACAGACAGAGAGAGAGGAGGAGG 134208
   148 yLeuProArgGluValPheGluGlnAspAlaSerGlyAspGluArgSer 165
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134207 GCTGTCAAAACAGACCTTTGAGTGTGGGGCCACGACGACACAGG 134158
   165 LyrProGluAspSerAsn..... 170
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
134157 ACGACACAGTAGCTAATGTCATCTGATCGAGTCTGATCCGAGTCCATC 134108
   170 ..... 170
134107 AGTTGATTTATTGACACACCGTATTAAATATCGATAATGTTGCCAG 134058
   170 ..... 170
134057 GAAGAAAGCATTCAGAGATTTTTCCTTTTAAATATGATAGTCTGG 134008
   170 ..... 170
134007 ACAAAAGACCAACACACACTGAATAAGCTGTCTCTGATGGAAAGAT 133958
   170 ..... 170
133957 AGAAAAAGACAGACTACAGAAATTGATATCTGTCTTATTCCTCA 133908
   170 ..... 170
133907 CATTCATATTATCATATTAAATAGTAATAAGTAAAGGGGTCTGTAT 133858
   170 ..... 170
133857 CATATTACTCTATATAAACAATCTCTGCAAGCATTAAGGGGCACTGCTC 133808
   170 ..... 170
133807 TCCTGGAAGAAATATAAACAACATGATCTCTCTTGTCTACCCCTCC 133758
   171 .....AspSerArgArgGlySerGlySerProGluAspT 182
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
133757 TTACGATCCACAGATGACGCAAAAGCGGCTCAGTCCGACGACGAA 133708
   182 hrProArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAla 198
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
133707 CTTCCTGAGGGAGCTGGGCTTCACGATCTTTGACCTGTAAGAACTCCACC 133658
   199 AlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaProGluAspVal 215
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
133657 CCAGATGCTCTCTCTGCGAGCTGTCTTACAGCAGCAGGCTCTGAGAGCAT 133608
   215 LAspArgArgAsnGluThrLeuArgArgGlnHisArgProProAlaLeuL 232
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
133607 GGACCAACGCAATACCGAGAGCGCCGACAGGGGCGCACAGTACCTGC 133558
   232 euThrLeuTyrProAlaProAspGlu..... 240
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
133557 TTGGGCTTTTACCACCCCTGATGAGGTTTCTCAAAAGCATTTTGTGATC 133508
   240 ..... 240
133507 TTGAATAATATGATTTTATATAAACAOTGCATCAGCAAAATCTTAAC 133458
   240 ..... 240
133457 AATATTTCTCAGCAAAACACACAAATTTAAATCAAAATTAATTAATTTGAT 133408
   240 ..... 240
```

```

133407 AGATTGTCACTGGAATATGCCGGTCTGACAGATGCAATGTGATTTA 133358
240 ..... 240
133357 AGCCTTGTTGTTATTTTGAATAGCTTACATTGACCTATGACATTTATG 133308
240 ..... 240
133307 CCACTGTTAAAAATGCATGGCAGCTTGAATGTTGTGATGTAGTGG 133258
240 ..... 240
133257 TATTGTTACGCTCTGGGTAAAGTAAATGTCCTTGTCTCAAAAGTCAGC 133208
240 ..... 240
133207 CATTACAGAAGATAGTTTCTGATTTGTCCTTGTCTTATACAGCTCT 133158
240 ..... 240
133157 GTTGACACAGTTCACTTTCTGTGGTCCAGCTTGCATGACATCACTGT 133108
241 ..... AspGluAlaValGlu 245
133107 AATATATTGTTCTTGCTGTGTGCCGCCGCGACAGAGCGGGGCTGTGAA 133058
246 ArgCysSerArgProGluProProArgGluHisPheGluGlnArgIleLe 262
133057 AGATGCTCCAGCTGTTGAAGTGCCCAAGAGCAGCTGTGCCAGAGATTAT 133008
262 uValLysCysLeuSerLeu..... 268
133007 GGTAAGTGTCTGTCTGTGAAGTGAGTAGTAAAAATATTAGATTATGT 132958
268 ..... 268
132957 CATTTATATGTCAAGCCCCCAATATTGTAACATTCTTCATACGCTTT 132908
269 Lys..PheGluIleGluIleGluProIlePheGluIleLeuAlaLeuTyr 284
132907 TCAAGTTGAAATGAAATGAGCCGATATTGGAATCCTTCCTGGTAT 132858
285 AspValArgGluLysLys..... 291
132857 GACATCAAGAAAAAGAAAGCATTTTTCCTCATTAGAGACTTGC 132808
291 ..... 291
132807 ACTTACGTAATGCTGAGCTCATGCTCCGCTAATTCAATTCCCTGTCTT 132758
291 ..... 291
132757 AAAAGGCTAAATGCTCTACAGTTAGATCACCAGTCAAAATGTAATT 132708
291 ..... 291
132707 AATTTCCCTGACTGGAATTCATGATGCTCATCTCTTCCTTCCTCCG 132658
292 ..... ILeSerGluAsnPheTyrPheAspLeuAsn 302
132657 CTTATCTTCCGCAACAGATCTCAGAGAAATTTCTACTTCGACCTGAAGT 132608
302 GluAspSerMetLysGluLeuLeuArgAlaHisGlyThrHisProAlaIle 318
132607 CAGATCAGATGAAGGGCTACTTAACACACACGCGCTCACACAGCCATT 132558
319 SerThrLeuAlaArgSerAlaIlePheSerValThrTyrProSerProAs 335
132557 TCTACCTGCGCCGCTCAGCTATTCTCCATCAGCAGTACCCCTGCGCA 132508
335 PhePheLeuValIleLys..... 341
|||||

```

```

132507 TATCTTCTGTGATTAAGTACTGAAGCTGCCACTTCACCTACATTC 132458
341 ..... 341
132457 TGCTTAATGTAATGCCATAAATATAGATCAGGTGTGTTGTGCAGACATGA 132408
341 ..... 341
132407 TCTAGTGTGCTAGATATACAAATGCAAAATATTGTTTGGCGCTAGCT 132358
342 ..... LeuGluLysValLeuGlnGlnGlyAsp 350
132357 TAACATGTTTGCATGCATTTTCAGCTGAAAAAGTCTCTCAACAGAGAT 132308
351 ILeSerGluCysCysGluProTyrMetValLeuLysGluValAspPheAl 367
132307 ATCAGTAGAGTGTCTGAGCCCTACATGTCATGAAAGAGTCTGATTCCT 132258
367 a.Lys..... 368
132257 AAGGTAAATGCAAAAAAATGTTACTGCTGTGATGTCATAACAGCCACC 132208
368 ..... 368
132207 AGTAACTTGACTGTGAGTCAGAACACATTAATTCATTCTGTGTCACAAA 132158
368 ..... 368
132157 GGTTCATTTTGTCTGTGTGGAACCTTTAATCTTCTGACAGCTACTCT 132108
368 ..... 368
132107 GAGGAAGAGACAGTCACACATTTGCTGTAAATAAAGAGAACAGATC 132058
368 ..... 368
132057 TGTGTAATAAAGAGTAATTAATCCACGCTGTGTGGGTGTGCGA 132008
368 ..... 368
132007 TGTTAGATTTCATTTACAGCTTTTGTGATTTACAGTTCACAAATCAGT 131958
369 ..... AsnLysGluLysLeuGluLysLeuArgLeuAlaGlu 381
131957 TCCTGTTTCAGCACAAAGAGAGCTGGAGAGCTGCTCAGCAGCAGAG 131908
382 GluPheCysThrArgLeuGluYargTyrArgMetProPheAlaIlePheAl 398
131907 CAGTCATGCAACCGCCTTGAGCGCTTTCGATGCTTTTGGCTGTGACAGC 131858
398 aValHisLeuAlaAsnIleValSerSerAlaGlyGlnLeuAspArg.... 413
131857 CATCCACCTGGTTACATTTGCAGCAGATAGGGCAGTCTTATGCGTACG 131808
414 ..... AspSerAspSer..... 417
131807 ACCCGAGCTTGTATTCGCGTACCGGTATTCAGCCACCTGTAAATGTCT 131758
418 ..... GluGlyValArgArgPro..... 423
131757 AATGCCCTTTACTGTCATGCGTGCTAAATATCTCCCTATATTAATGCT 131708
424 ..... AlaTyrPheAspArgArgArgGlyP 433
131707 AACACAGAAAAAAGGCCATGGGACATGGAATGAGACGAAAGAAAGGGGG 131659
433 roGlnAspArgAlaSerSerGlyAspAspAlaCysSerPheSerGlyPhe 449
131658 ..TTTGACCGATGATGATTTGCGGACGACACATGTCAATTTTGCACCTTC 131611
450 ArgProAlaThrLeuThrValThrAsnPheLysGln..... 462
131610 GGTCCAGCAACATTTGACCGGTACCAACTTCTTCAAGCAGAGTGTGACTGTC 131561

```

462 462
131560 TGTCAATAGTAAAGCTTGAGTGTGGCTATATTTGATTTGTTGCTATC 131511
462 462
131510 ATCAGCTATACATACAAACAAATGATGATTATCTTAAGACACACATA 131461
462 462
131460 GACTGTGATCAAGTATGTATTGCAACTCAAGTTATAGGAAAAAAGT 131411
463 Glu..AlaGluArgLeuSerAspGluAs 471
131410 TAAATTTCTTGTGTTTCATGAAAGGAGGGGACCGAGCTAAGTACAGAGGA 131361
471 pLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArg 488
131360 CCTTTATAGTTTCTGGCAGATATGGCAGACCGCTCTCTGTTGCGGAC 131311
488 rLeuArgProValThr 493
131310 GACTGAGGCTGTGCACAGGTAAAGCAATTACTTGTACAAAGATA 131261
493 493
131260 ATTGTTCTAAGATGTATGATGTTGTTGCTCTCCCTAACTTAAGATGC 131211
493 493
131210 ACGTTAAAGGTGTATCCATTTTCTTCTTGTGCTGTGATTTAT 131161
494 AlaGluLeuLysIleAspIleSerProAlaProGluAsn 506
131160 TTGCACCTGCAGCCAGTTGAAGATTGATTCCTCCAGCTCGGAGCTGG 131111
507 ProHisPheCysLeuSerProGluLeuHisIleLysProIleProAs 523
131110 CCTCATTTACTGCTGTGCACAGAACTGCTTCATGTAAMCCCTACCCACA 131061
523 pProArgGlyArgProThrLysGluIleLeuGluPheProAlaArgGluVal 540
131060 CCCACGGGTGGCCCAACCAAGAGTGGTGGAGTCCCTGCTGCTTATG 131011
540 aLysrAlaProHisThrSerTyr 547
131010 TGTACACGCCACATACCTCTATAGTGGAGGCCCCCAACTCACAGAA 130961
547 547
130960 TCATTTATTTCTCTCTGTCAATATCAATTAATTTATTTGTCCTTAT 130911
547 547
130910 TTGCTTATTAATCTGATATAATTAAGAGATAAAAACACAGTAATACCT 130861
547 547
130860 ATTGAACAATATTTTATATACATTTCTTAATAGAAATACATTTCTCAG 130811
547 547
130810 TAAATTCATGTTTACTCAAGTTTAAATATGTAACAATTTGTTCAAGTAA 130761
547 547
130760 TAAAGCTAAGTTCACAAGTACAGACAACTCCCATCATACAGGTCAAA 130711
547 547
130710 CGATTAAAGAACTTTGTAGTAAAGTCAATAGTCAAGTTCAGACTTACGTACA 130661

547 547
130660 TGTACGGAGAGAAAGAGATGCAGTCAATGTTGCAGTAAAAAAACCTG 130611
547 547
130610 TGCCGAATCTTAATGGCCAATGCTTGCCCTGCTCTTGCAGATGCMAA 130561
547 547
130560 GAGAAACAAAGATACACACATTAGAGGATATGAATTAACACGGGGCGTT 130511
547 547
130510 TTAACAATAATCTTACCGTTTCTGTAGTTAGTGTAGCTTTCACCTG 130461
547 547
130460 ACTCTGATGCTCTGCTTTTGAACCTGAATCAGTTCTTTAGTCTCTCA 130411
547 547
130410 AAAGTCACTGTGCTCTGCTCTAGTCTAGCTAGACGGGAGAACAGTTTC 130361
547 547
130360 CTATATCATTTTTTGTGTGTATGCTGATTCCTCTTAGTATTAATTA 130311
547 547
130310 CAGAGGAATTTCTTTACTGCAGCTCTTTGATGCTGCTGGTGGGTACATA 130261
547 547
130260 CTTATTGAGATGTGTCACAGCTCTTTACATTTGTCCTGAGTCCACTT 130211
547 547
130210 CTTCTGCTCCAGTTTGGCAGCCTGTCTTTCGTACCTCGGCTGCAGCC 130161
547 547
130160 TGTTCCTGGCCATCTGGGTCTGTTGAAGAAGCATCTGTCCATGTTATTC 130111
547 547
130110 CTCTACAGCAGCAATCTCACGGCAGTGGCAGCGGCTGACCCCGTAT 130061
547 547
130060 AACATGTGGCTTTCTAATGACACACTCTCAGATGAAGGGTCTCTATAT 130011
547 547
130010 CTGCAACATCATGAAACCAATGTACAGCAAAATACCAAAAGTCTTA 129961
547 547
129960 GAAAAACAATCTTATCCCAAGAAACAGTGTCTTATTCAGATTTTGGAG 129911
547 547
129910 ATATGACCTGTGTTTCTGGCCCTCTGCATGTTTTCGTTGCTGCAGTTT 129861
547 547
129860 GATTACAGTGTGTCTCTCTCTCTCTAGACAGACGCCCTTCAAGTAG 129811
547 547
129810 CCAGCAGACATGTTGTGACATATTTCTGCCATCACTTACCTGTAGCAA 129761
547 547

129760 CATTATACTGCTTAGCCCTGTTTTCATCTCTGACGCTTCTCAGCAT 129711
547 547
129710 GTCTCCCAAGCCTTGAGATATCCTTTCATGGTTCTCAAAAGTCTGGCT 129661
547 547
129660 ATTCTGTTATCTTGGCCCTGGGTGTTTCAGAAATGTAATTTAGGAGATC 129611
547 547
129610 GCCGAAGTATGAATCCAAAGCTGACAGCTTGACCTGCTCTGGTG 129561
547 547
129560 AAACATTGTTTTCTGCTGTGTTCTCCTCAGTGCAGCTGCAGGGTGTCT 129511
547 547
129510 CTTAAACCCCAAGCCAAACCATCAGACAGACTCCAGATCAGATTTCAT 129461
547 547
129460 AGTATTATTCCTCCTACTGTAGTCCAGCTGTTGACAGTTCTGTGTTTC 129411
547 547
129410 TTCTCAGATTGTATGTCAGAAAAGTTAGTAAATTCGTAATAGAT 129361
547 547
129360 CCCATCTGTTTAAGATCCAGTGTGTGTGGCTGTGTGAATTTGGAGG 129311
547 547
129310 GTAGTCAGTGCACTTATATTTGGATGTATATGATAGTAATAATTA 129261
547 547
129260 GCTCAGCTGTGATGTCATCAGTGCACCTTACTCTTTGACCTTGCCCTTA 129211
547 547
129210 ATGTTTAACCCCTGAGATCTGATCTTGACATTTAGGTCACAAACATG 129161
547 547
129160 TGACACATTACTGATATATTTTGGTTTGACACAGATCTTGGTGTGTT 129111
547 547
129110 TCATGCCACACTGCTGTGTAATGTGTAAAAAGATTAATTCATAGTCTC 129061
547 547
129060 AAACATCTTAATCAATACAAAGAAATTAATCAGCTGTCCAAAGTAC 129011
547 547
129010 CTCAGATGTTGACTTTAGTTTGTGAAGTTACGGTACTTTGACTCCACA 128961
547 547
128960 ACATCCAAATGCTGGTGCACATCCTGTTACATAGATATTTGCATTCAT 128911
547 547
128910 GAAATCTTTCATTTACTCTTCTAGATTTAGATGACAGTTGGTA 128861
547 547

128860 GATGCTTTGACAAACAGAAATGTGCCAACTTTGCATTCATATTTAAAT 128811
547 547
128810 TGGCAACAAGATCTTCCCTTCAAAATCAGCATTAACAAAAATATATGTA 128761
548 Arg, AsnLeuLeuTyrValTyrProHisSerLeuAs 559
128760 AAACCATCTCATTTGCAGAGACTGCTATGTTACCCAAAGACCTTGAA 128711
559 nPheSerSerArgGlnGlnSerValArgAsnLeuAlaValArgValGlnT 576
128710 CTTCAGTAGTGTCTCAGGAGCTCAGTGAAGACATCGCTGTCAAAATCCAGT 128661
576 yrmEtrhGlyGluAspProSerGlnAlaLeu 586
128660 ACATGCGAGAGAGAGACCCCAAGCTTTCAGGTGAGAGAGAATGCAA 128611
586 586
128610 GGACTACCAAGTATAAATACAAATGTTCTAAGAAAAACAGCCGAAT 128561
586 586
128560 CTGTATTCTGAGAAATTTACCTGGCTTTAGTAAATCTGACAGTAAGTAATA 128511
587 Pro, ValIleP 590
128510 AAACATCCCTTAATGATTTTGGTATTTTATTTTTCACATAGGTCATTT 128461
590 heGlyLysSerSerCysSerGlnPheThrArgGluAlaPheThrProVal 606
128460 TTGGAGCTCAAGTTGTGTGATGTCACAAATGAGGATACAGTCCATC 128411
607 ValTyrHisAsn 610
128410 ATCTACCAACAACAAGTAGGCTTTCAGACATTAACTGCTGTAACTCC 128361
610 610
128360 TTTGTGTTTCCCTTAATTCGTGTGTTTAACCTCTTTTGTGTTAAAAA 128311
611 ..Lys..SerProGluPheTyrGlnGluPheLysLeuHisLeuProAlaC 626
128310 AAAAAGTCCCTGAGTGTATGAGGAGATGAAGATGAAATTTCTCTGCCA 128261
626 ysvAlThrGluAsnHisLeuLeuPheThrPheTyrHisValSerCys 642
128260 ATCTGATGACATACACCACTCCTGTTCACTTCTACACATCAGCTGTC 128211
643 GlnProAlaGlyThrAlaLeuGluThrProValGlyPhe 656
128210 CAGCCCAAAACAACACATCTCTAGAGACTCTGTGGCTACACTGTGAG 128161
656 656
128160 TTCACCTAGTCAGTCTATCTTATCTCCACCCACAGCCCAAGTTT 128111
656 656
128110 GATAATCTTCTGTTGTCATGCCATCTTAAAGATCTCGTCTTTCTT 128061
657Thr..TyrIleProLeuLeuGlnHisGlyArgLeuArgThrGly 670
128060 TCGACTGTAGTGAATCCCTGTGATGCGAGCGGCGCAGTGAAGACAGCG 128011
671 PropheCysLeuProValSerValAspGlnProProProSerTyrSerVa 687
128010 TCCTTCAATTTACCTGTCTCAGTGAAGAGCCGCCACGAGCTACTCTGT 127961
687 IleuThrProAspVal, Ala 693
127960 TCTCACCCCTGATGTAAACAAGCACTGGGGAATATCAGTTTGCCTGAATA 127911

```
693 ..... 693
127910 GACCCTTGGCGTTTGTACTGCGTGTAAATGAGATGGATTTGCGTGGAT 127861
694 ..... LeuProGlyMetArgTTPValAspGlyHisLysGlyValP 707
127860 GTAGGTCACGCTTCCAGGATGAGATGGGTGATTAATCAACAAAGGAGTGT 127811
707 heserValGluLeuThrAlaValSerSerValHis..ProGln..... 720
|||||.....
127810 TTAAGTTGAAGTACACAGACGCTTCTCGGTTCATCTACGATCAGTACACGAC 127761
720 ..... 720
127760 GCTAAATATTGTGCTTTCATCTCATCATCTGACACGACATCATTTCAAG 127711
720 ..... 720
127710 CATCCATTGATCTTAATGTTAATAGTAATTAATAGTACTGTAAC 127661
721 ..... AspProTyrLeuAspLysPhePheThrLeuValHisValLeuGlu 736
|||||.....
127660 GTAGGACCCCTCTCGATAGTTCCTGACGCTGCTGTATGCTCTGGAAG 127611
736 LuGlyAlaPheProPheArgLeuLysAspThrValLeuSerGluGlyAsn 752
|| .....
127610 AGTACTCTCCCTCCGCTCCGCTGAGGATGTGATCATTTACTGATCAAC 127561
753 ValGluGlnLeuArgAlaSerLeuAlaLeuArgLeuAlaSerP 769
|||||.....
127560 ATGGAGGACAGTTAAAGACACAGCATGTGGCTGACATGAGGGCCCTGTT 127511
769 ogLuProLeuValAlaPheSerHisValLeuAspLysLeuValArg 786
|||.....
127510 GGATACCTGTGTGACGCTTCTCCACACAGCTGCTCAACAGATCATTTCAAC 127461
786 euValIleArgProProIleIleSerGlyGlnIle..Val..... 798
|||||.....
127460 TCATCGTTTTCACCGGTCATCGAGGCCAGCATGTGTAATGAGCTGTG 127411
798 ..... 798
127410 AAGCACCCTGCTGCTGTCTAGTCTACAAATGTGCAATTTGGAGTGTCT 127361
798 ..... 798
127360 CTGTGAACAGGCGTGAAGGCTCATTTCTAATAGTGGGTTAATAGTTGAT 127311
798 ..... 798
127310 GGTAATGATCAAAATTTACTGTGAGAAATATTTGACATTTTGTCTCA 127261
798 ..... 798
127260 TACACATAAAAACACTGACTTTCATGCAATGCTAAACATGTTGTTTAA 127211
799 ..... AsnLeuGlyArg 802
|||||.....
127210 AATAATCCTACATCTTGATTTCCAACTCCTGCACGTCAACCTTGGCAGG 127161
803 GlyAlaPheGluAlaMetAlaHisValIleSerLeuValHisArgSerLe 819
|||||.....
127160 ACTGCTTTTGAACATATGGCTTACTGTGTAACACACACACAAACAACT 127111
819 uGluAlaIleAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaIleTyr 836
|||||.....
127110 GGAGGGAACAGGACCATATATGTGTGCAACAACCTGTGGCTTC..TACA 127062
836 alHisTyrAlaPheArgLeuProGlyThrGluProSerLeuProAsp... 851
|||||.....
127061 TCCAATACGTGCTTTCGCTGCGCCACATGCTGAACATCAGTGCGCCCAACA 127012
851 ..... 851
127012 GGTGAGACCTCGCACCACTGTGCGACAGACAGATTTATTTGTTGATTTGCT 126962
851 ..... 851
126962 GGCAGTGTCTGTAGACCGGCGCGTTGATCTGCCCTTATCTTCTCTGTGG 126912
852 ..... GlyAlaPro..ProValThr 858
|||||.....
126911 CTCCTATCCAGCTTTGAGAGAACCCAGTTCACAGTCCCATAGATGCCG 126862
858 alGlnAlaAlaThrLeuAlaArgGlySerGlyArgProAlaSerLeuTyr 874
|||||.....
126862 TGCAGTACGCAACCATATCGAGGGCACTGGACGCCCAAGCAGCTGCAC 126812
875 LeuAlaArgSerLysSerTLeuSerSerSerAsnProAspLeuAlaValAl 891
|||||.....
126811 CTGTCCGCTCCAGAGCATCAGCAACTCCAAACCAGACCTGGCCAGTTTC 126762
891 aProGlySerValAspAspGluValSerArgIleLeuAlaSerLysLeu. 907
|||||.....
126762 ACCAAGTTTCTCCGATGAAGGTGCAAGATCTATGGAAGCAAGGTGA 126712
908 ..... 908
126711 GCTCTTTCGGCTATGAAGCTTTTGTCTTTATTTACAGCTTAGAGAGTG 126662
908 ..... 908
126662 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGA 126612
908 ..... 908
126611 ACATTTATTCACAGTGAATCTGTTACATCTCTGTAGAAAGTGAATTTTC 126562
908 ..... 908
126562 TCTCTGTTTCAACAAACACAGCTTTTACAAACAAAGTAAAGTTT 126512
908 ..... 908
126511 GGGGTATTTCTTGCCTCCCAAAAACTGAAGTTGACCTTCTGGGCTA 126462
908 ..... 908
126462 GGGCTTCCCAACACAGTGTCTACATGTGGCGTGGCCCTTGTGTGAGA 126412
908 ..... 908
126411 CTCTGTCTCTCTGGGCAATATTCATCCACCAATATTTGCCATTCGCTC 126362
908 ..... 908
126362 TTCTGTATATATATATATATCTTTATTTGGGTAAAGTGTGAGAACGAGCT 126312
908 ..... 908
126311 GGAACAACTCCCATATCAGGAATGACATCATTTAGTTGACCGCGCCGT 126262
908 ..... 908
126262 GGAGCTGATCTCATTTTCTCGCTCTTTAGAGTATGTGTATATATGTAT 126212
908 ..... 908
126211 TTTTATGAAGATCATTTATGCTTTGTGTTTGAACATGTCTGTGAGC 126162
908 ..... 908
126162 CCTTGATTTCTGTGCCAGTGTGAGATTTCTGTGTGTTTAAGCTTAT 126112
908 ..... 908
```

126111 TCCTGGCCCCGCCCCCTTCTTGTGTTTTCACATTTATCTCC 126062
908
126061 AAATTAATTACTGTGTGTATTAATAGCTCCCTCCTCCTCCTG 126012
908
126011 CCTCGGCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 125962
908
125961 CATGACCGGCTCCACTCTGTGGTTACTGTGTGTGTGTGTGTGT 125912
908
125911 CCAGATCATTTGCTTGGCGGAAACCCCACTCCAGCTGTGATCG 125862
908
125861 GTGGCAATGCCCTCTCTCCGCTCTCTCTCTCTCTCTCTCTCT 125812
908
125811 CTAAACAACACTCTGTAGTCATTTGTGTGTGTGTGTGTGTGT 125762
908
125761 TTTCAATGCTGCACCTCTCTTATCTGTGAATACTACTATCT 125712
908
125711 CATTCATCTCTCTCAGACACCATAAGCCACACCTCTGTGTGT 125662
908
125661 GCCTGTGTGAGATTTTCTACTAGCTCCTCCTCCTCCTCCTC 125612
908
125611 TGGTACCACCAATTTTCCATTTCACTATTTTGTGTGTGTG 125562
908
125561 TGTTCACCTCACCATCAGACATGTGACAGAGAGCCGAGTCG 125512
908
125511 GCATGTCTGTCTTCTGTGACAGATGGCTTGTTCAGTTCACAG 125462
908
125461 CACATTTGCCAAGAGTAAGTGTCTCTCTCTCTCTCTCTCTCT 125412
908
125411 GTTCAACTCAAGAGATGATTTGTGTGTGTGTGTGTGTGTGT 125362
908
125361 CTGTATCTCATTCATCTGATCAGATCCTCATATGTAGTTAAAT 125312
908
125311 CACTGATCTAAACATTTTAACCAAGTATGATTTATTTATTTAT 125262
908
125261 CATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 125212
908
908 908

125211 AATGTTTAGATGTTTCTGTTTCAATTTAAATCTAATTCATAT 125162
908
125161 AACATTATATTAATGACAAAATTTCTAGTTAAGCATGATGTGA 125112
908
125111 TTTCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 125062
908
125061 GGAATATTTAGATTTTGTAAATATATGACTGTCTGATTTG 125012
908
125011 AATACCTTTAGAGTGTGCAAAACCTTTTGGAAATGCTCTAC 124962
908
124961 AATGGTTAATTAATTAATCTGTATTAAGGAACTTTGTAAC 124912
908
124911 CAGTAAGGTGGCTAAATTTATGCTTTCTAGGAATCATAGTAA 124862
908
124861 CAGCAAAACATTTAATGATGACAGAGATTAATTAAGTAC 124812
908
124811 ATGTTTTCACACGTCGCATTTGACCAAAAGTATGATATTAAC 124762
909
124761 TGACTTATGTTGTGCTCTGCTGTGAGTTGCTTCAATGAG 124712
914 eugIntpValIvalSerSerAlaValArgIuValIleuGlnHis 930
124711 TGCAGTGGGTGTCAGACGCCACGAGGAGGAGGAGGAGGAG 124662
931 AlatrPhePhePheGlnLeuMetVal..... 939
124661 GCCGTGTTTCTCTCCAGCTGATGAGAGCATGCTGTGTGTGT 124612
939
124611 AATCCATTAACACATTTTAAATTAATGATCTGAAAGTCCAT 124562
940
124561 TTTCCCTTTCTTTTACTCTTTTGTGACTCTTCCCGACCAAGAG 124512
942 etAlaLeuHisLeuLeuGlnArgLeuAspThrProArgLysLeu 958
124511 TGTCTCATCACTTATTTCTGACCTCAAAAGTGTATGACGAA 124462
959 ArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGly 975
124461 CGATTTCCAGACCGCTTTGTGTGTGTGTGTGTGTGTGTGTGT 124412
975 IGLyLeuGluValIleThrArgValHisLys..... 985
124411 CAGTGCAGATATTTCTGTGTGTGTGTGTGTGTGTGTGTGTGT 124362
985
124361 TTTCTCTCAAAAGTGTGAGTTTGTGTGTGTGTGTGTGTGTGT 124312
985
124311 AACATAATGATGCTGCTTAATATGTGAACATCTCTTAAACAT 124262

```
986 .....Asp 987
124261 TGCCCTATTTCACATGACATCCACACGTTTTCCTCCCTGTCACAGATG 124212
987 aGluleuAlaGluHisLeuAsnAlaSerLeuAlaPhePheLeuSerAsp 1003
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124211 TTGAACCTGTGGAGAGTTAAATAGCAGTGGCCCTCTCTCGTAATGAC 124162
1004 LeuLeuSerLeuValAspArgGlyPheValPheSerLeuValArgAlaH 1020
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124161 CTGCTTCTCTCATGACGACGAGGCTTTGTTTCACTCATCCGCTGTA 124112
1020 sTyrLysGlnValAlaThrArgLeuGlnSerSerPro..... 1032
124111 CTACAAACAGAGTGGTGGTCTT.....TCCTCTCCCTGTTTGTCTTAT 124069
1032 ..... 1032
124068 TATCAAAACATTGCTTTCACATATATTGTTCTCAGCTGCTGTGGCTT 124019
1033 .....AsnProAlaAlaLeuLe 1038
124018 TAACGATTGGCAACAAAGCTTCACACGCGCAAGAAATCCAGCTCCCTGA 123969
1038 uThrLeuArgMetGluPheThrArgGlyLeuCysSerHisGluHisTyrV 1055
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123968 TGCCCTGAGAGATGGACTTATTCGATGCGTGTAGCCATGAGCACTACG 123919
1055 aThrLeuAsnLeuProCysProLeuSerProProAlaSerProSer 1071
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123918 TCATCTCAACCTGGCCGTGTCTCAGCTCCAGCCCTCCAGCCTGGCCCTCA 123869
1072 ProSerValSerSerThrThrSer.Gln..... 1080
123868 CTTTCACCTCTCTCCACACGCTCACAGGTGCTATTACATGCTGGAGGCT 123819
1080 ..... 1080
123818 TAGTACAATCATTTACATTCCTCATCATATCCAAATATCTTTTCTTCA 123769
1081 .....SerSerThrPheSerSerG 1087
123768 TTTTACATTTCTTATTTTACGTTCCACAGAGTTTCAGCATTTTCCAGCAT 123719
1087 naLaProAspProLysValThrSerMetPheGluLeuSerGlyProPheA 1104
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123718 GGTGCAGAGATCAAGGAGTGGCCACAATGTTGAGAGCTCTCCGTCCTTC 123669
1104 rGgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAlaLeuAla 1120
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123668 GGCATCAACATTTCTCTCTGCTGCTGCTGTTATCAGAGCTCTCTCTATC 123619
1121 LeuGluProGluAlaGly..... 1127
123618 CTGTGCTCTGATGTGAGGAGGTAGGAGTGAATCTTCTTTCTTTT 123569
1127 ..... 1127
123568 ATGCTGGCTATTGGACCATGCTAATGTACTGAAAAATGCTGTATG 123519
1127 ..... 1127
123518 CTTCAGTCCAGTACATGAAAACTGAGTGTGTGTTTTCCTACACACAG 123469
1128 AlaPheLeu.LeuHisLysLysAlaAlaSerAlaValHisSerLeuLeu 1144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123468 CATTTCTCTCTCCATTAAGCAATTAAGCTGTACACTCCCTGTGT 123419
1144 ySgLYHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArg 1160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123418 GCAGGCATGATGACATCTCTGCTACAAAGACCTCAGTCAAGGCCCAT 123369
1161 VALAlaGluLeuTyrLeuProLeuLeuSerIleAlaAspThrLeuPr 1177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123368 GTTGCTGAGCTTTACCTGCCCCCTCATTCCAATTGTCTATGAGAGCACTGAA 123319
1177 cArgLeuHisAspPheAlaGlyProGlyGlnArgSer..... 1190
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123318 TCAGCTTCATACCTTCTCT.....GGTCAGTGTACACCCCCACACA 123278
1191 .....Arg.LeuAl 1193
123277 TTTTCATACCTTCGCAAAATGTCACCTTCGTGGCATTTTAGATGTTTAT 123228
1193 aSerMetLeu..... 1196
123227 TTCAATGCTGTGTGATATGTTTGTGTAGACTCCCTCTCGGGTCC 123178
1197 ..AspSerAspThrGluGlyGlyAspIle.....AlaGlyThrIle 1210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123177 GGGATGCCCTGACCCCATGACGATGCTGACCTGGATAGTGGCAACACATC 123128
1211 AsnProSerValAlaMetAlaIleAlaGlyProLeu..... 1223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123127 AGTCAGTCTGTGGCATGGCAATGGCCGCTCCCTTTACACATGCCAA 123078
1223 ..... 1223
123077 AGCCACATCCTTTCGCTGCTACCCACAGCGGTGAGTGAATAGGCTGGGGTC 123028
1224 .....Ala..ProGlySerArgAlaSerIleSerGlnGlyPro..... 1235
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
123027 AGGTTGTTCCCTGCGCATTAACAGAAACAGTCAAGATCCACCTGTG 122978
1235 ..... 1235
122977 TGGGAGACTGTGGCTGACCTGACAGAACACACACACACTGGCCTTCTGTA 122928
1236 ..ProThrAlaSerArgAlaGlyCysAla..... 1244
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122927 ACTCCACTAAAAAACAACAAAATGTGCATCTTTCCACATTGTCTATC 122878
1244 ..... 1244
122877 TTCCTCAAGATAACTGCTTGTGTGCTATAGTGGCTCTGCTGCCCT 122828
1244 ..... 1244
122827 CTCTATAAACHTCTTTCATGTTTGACATGAAGGCCAACTTACCGTG 122778
1245 .....LeuSerAlaG 1248
122777 TGTGTTGCTCTTGCAGCCCTGGCCGCGCATCCAGCTCTGTGTCACACA 122728
1248 uSerSerArgThrLeuLeuAlaCysValLeuTTPValLeuLysAsnThrG 1265
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122727 GTGAGCCGAGCTGCTGCTGTGTCTCTCTGCTGGTGTGTAACAAAGCAG 122678
1265 lUProAlaLeuLeuGlnArgTyrAlaThrAspLeuThrLeuProGlnLeu 1281
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122677 ATGCAGCTCTCTGAGCGCTGGGTGTCTGATTTTGACAGTGCACAAATC 122628
1282 GlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheGluTyrLy 1298
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122627 AACGCGCTGTGATCTGTGTCAGACTCTGCATCTCGTGTGTAATATA 122578
1298 s..... 1298
122577 AGTAAAGTACTTGTGATTTTGTCTTCTTTTTCACCTGCTTAGCTGT 122528
1299 .....GlyLysLysAlaPheGluArg 1305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
122527 TGACATGTTTACTTTTGTATTAATAGGGAAGGCTCTGAGAGG 122478
1306 IleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuG 1322
```

```
|||||
122477 AFTAAATACCGTACGACATTTAAAAATCCCTGGACATGAAAGCAGCGGTGGA 122428
1332 uGUAlAlleuGlyThrIleGlyAlaArgGlnGluMetValArg. 1338
122427 GGAGCCACTACTGGCAACCATTTGGGGGCTGCTCAGAGATGGTTGTGTGTT 122378
1338 ..... 1338
122377 GCAGAGTAATACGTCTTCTGCTGCTGTGTGTGTGTGTGTGTGTG 122328
1338 ..... 1338
122327 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 122278
1339 ..... Ser ArgGluArgSerProGlyAsnProGluAsnValArg 1352
122277 TCTCAAAATCTTTTAGAGAGAGTCCCTATAGCAGTCAAGAAATGTGAGG 122228
1353 TPArgLysSerValThrHisTrpLysGluThrSerAspArgValAspLys 1369
122227 TCGAAGAAAGACGTCACTGCTGGCGCAAAATACAGACAGAGTCGACAA 122178
1369 s..Thr..... 1370
122177 GTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAA 122128
1370 ..... 1370
122127 AGATGTTTCATGTGCGTATAGCAACTAGTGTTCATCATCAGCGGTGCG 122078
1370 ..... 1370
122077 GATCAATCTGTCTTTAAATTTTCCAAGTCTGTAATGCTGANTGTAAGC 122028
1371 ..... LysAspGluMetGluHisGluAl 1378
122027 TAAGCAATTAATTAATCTTCTTAGAGCTAAAGCTGAGATGGAGCGAGGTC 121978
1378 aLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuValValLeuAspT 1395
121977 CGTGTGTGAGGAACCTTATCAACCGAGGCTTCTTAATGCTGCTTGACA 121928
1395 hrLeuGluIleLeuVal.Gln..... 1401
121927 CGCTGGAGATTTGATGCTAGAGTACAAATGCTTCAAGCTGTGAAGTGC 121878
1401 ..... 1401
121877 GAAATATTATGAGATTAATTTCTGTCTTCTTGTCTTGTGATTTTTC 121828
1402 ..ThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLeu 1417
121827 AGACTGTGTGGATCAGACTGAAAGAGTGTCTGAGTGGAGTAATG 121778
1418 LysValValLeuThrSerLeuGlySerAlaGlnSerAlaLeuThrLeuG 1434
121777 AGAGTCTCTCTCCACAGCATGGCAGGAACCAAGAGTCCCTTCTCTGCA 121728
1434 nHisGlyLeuAlaThrGlnArgAlaLeuVal..... 1444
121727 GCACCTGCTTCAACAAAGAGCTCTAGTTTTCAGGTACATAAATTAAT 121678
1444 ..... 1444
121677 TTTTGTGTAAATTTGAGATTAATTTCTTCTTATCTTTTACATGTT 121628
1444 ..... 1444
121627 GCAGCATATTACATCATAGTTAGATCATCATATAGATAGACA 121578
1445 ..... Ser..Lys..... 1446
|||||
```

```
121577 TGTGTGACCACCTTGAAGTGAAGCAACTGAATATACATGTGTGT 121528
1447 ..... PheProGluLeuLeuPheGluGluAspT 1456
121527 TTAATATACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 121478
1456 hrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysGlySer 1472
121477 CAGAACTTGTGCGGACCTATGCTGTGCTGTGTGTGTGTGTGTGTGTGT 121428
1473 ArgIleSerThrIleArgThrHisAlaSerAlaSerLeuThrLeuLeuMe 1489
121427 AGCGTATGTTCTGTGCAGAAAGTCAGCGCTGCTGCTCTTTATCTGTCT 121378
1489 TarGlnAsnPheGluIleGlyHis..... 1497
121377 GACACAGAACTTTGACATTTGGAATGTAGTAATATTAATCTTCATGTTC 121328
1497 ..... 1497
121327 TGATGTAATATTAATCTTCAATATGTTCATTAATCTTATAGGAAGTTTA 121278
1497 ..... 1497
121277 CTAGACATTTACTTAAAGCTGAGACAAAGAAAGAAATTTACATCCCTT 121228
1498 ..... AsnPheAlaArgValLysMetGluValThrMetS 1509
121227 TTAATTCCTTTTGACAGAACTTGTCTGGGTAAAGATGCAGGTCACATCT 121178
1509 erLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGluHisLeu 1525
121177 CACTGTCAATCACTGTGTGGGAGACATCAAGAACTCAATGAGGAAATCTTT 121128
1526 ArgArgSerLeuLysThrIleLeuThrTrpAlaGluGluAspMetGlyLe 1542
121127 CGACACTGCTTGAAGAACTATCTTGACATATGCTGTAAGATGATGTGAAT 121078
1542 uArgAspSerThrPheAlaGluGln..Val..... 1551
121077 AGGTGACACACCTTCCCGAGCAGGATTAATGAGCGAGTGTAGAGAT 121028
1551 ..... 1551
121027 CCCATTTGTAAGATGATGATGATGATGATTTTAATTCATTCCTGTGT 120978
1552 ..... GlnAspLeuMetPheAsnLeuHisMetIleLeuThr 1563
120977 TGAACACTGAGTCCAGATTTGTGTTTCAATTTGCAATGATTTCTCAC 120928
1564 AspThrValLysMetLysGluHisGlnGluAspProGluMetLeuLeas 1580
120927 GATACGTGTAATAATGAAGAACACAGCAGATCCAGAGATGCTTATGA 120878
1580 pleuMetLysArg..Ile..... 1585
120877 CCTGATGTACAGTAATGCGATGACATTAATCAAAATATAGTATCA 120828
1586 ..... AlaArgLysTrpGln 1591
120827 GACTTAATATTTGCTTTTGTAAATACGATTTGCAAGGAGCTATTCANA 120778
1591 LysSerProAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHis 1607
120777 ACTTCCCGACCTGCGTCTGACTTGCTCCAGAACATGCGCAAGAAACAC 120728
1608 AlaGluLeuGlyAsnHisAlaGluAlaAlaGluIncysMetValHisAlaAl 1624
120727 TCTGAGAGAGAAACCATGCTGAAGCTGCCCATGCTGCTGCAAGTGC 120678
1624 aAlaLeuValAlaGluTrpLysLeuAlaLeuGluGluAspGlnArgHisLeuP 1641
120677 AGCACTGTTGCAGAATACCTCAACATGCTGAGAGATTTGTGCTACTTCC 120628
```

1641 rovalGlyCysValSer..... 1646
|||||
120627 CAATCGGTTCGTCACTTTCAAGTGGGTGAGAACGACTGACTTAAGT 120578
1646 1646
120577 ATGTGATGCAGCAAGTAGTGGGTGTATATAAATGCATCATACAT 120528
1647 PheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSerA 1662
|||||
120527 ACAGTTTCACAAACATTTTCATCCAAACGCTCGGAGAGATCTGCTGTCTG 120478
1662 spAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPhe 1678
|||||
120477 ATGACGCTGTCTCCGAGAGAAAGAGGATCTGTCTGGAAAGTACTTC 120428
1679 ThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaIleGlyTyrPheTh 1695
|||||
120427 AGCAGAGCTGCTGCTAGTGGGCTCTTGAGCAAGCAGCTGCTCTTCA 120378
1695 rMet..... 1696
120377 CATGTACTGTCTACATTTGTCTACACATGTGACAGACTGCTGTGTG 120328
1696 1696
120327 TCCCATGGGCTGTGGGGTTTATATGATGTGGCGTGTGTGTAAACTGG 120278
1697 Gly.GlyLeuTyrGluAlaValAsnG 1705
120277 CCTACCAATGTGTGTGTAGCTGCAGGCTGCCATGTATGAAGCCATTAATG 120228
1705 LuValTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLys 1721
|||||
120227 AGGTTTCAAGATTCCTCTGCCCATCCATGAAAGCAACAGACTTTAAG 120178
1722 LysLeuAlaAlaValHisGlyLysLeuGluGlnAlaPheThrLysIleLeu 1738
|||||
120177 AAGTGGCTACAGTCATGGGAAGCTCAGGAAGCTTTAAACAAGTCTA 120128
1738 His..... 1739
120127 CAACCAAGTAGTTTGAAGGCTGATGACAGACAGATATCATTTTACTCT 120078
1740 GlnSerSer 1742
120077 GAAGTTCCTCAATGTTTGAATTTCTTTGTTTTCACCTTTCAGAGTCTA 120028
1743 GlyTyr.Glu..... 1745
120027 GGATGGAGGTAATGCTTATTTTAAAAAATGACTGATGAGCACTTACT 119978
1745 1745
119977 TTATTTATGTACATTAACATTCCTCTTCCCTTCCCTTATCTCCCCCT 119928
1745 1745
119927 TAATCCTTTGATGACAAATATATGGTGATTTTAAACTATTTTAA 119878
1745 1745
119877 TCCTAAAAATATGATTTTGTCTTAAAGCTGTGATTTAAATCATCAT 119828
1746 ArgValPheGlyThrTyrPheArg 1753
119827 CTATACATCCCATCTTCAAGTGTAGTGAAGATGTGGGACCTATTTCCGG 119778
1754 ValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheVal 1770
|||||
119777 GTGGGTTTCTTGTGCTCTCGTTTGGAGACCTGACAGACAGAGTTGT 119728

1770 ITyrLysGluProSerIlePheThrLysLeuAlaGluIleSerHisArgLeuG 1787
|||||
119727 CTACAAAGAGGCATTCATCATAAATGCGACAGATCTCTCACAGACTTG 119678
1787 Lu..... 1787
119677 AGGCATGTTGACTCACCGCAGCCTCAAGAAAAAGATTAATACCGCTTCC 119628
1788 GluPheTyr 1790
119627 CAATTAATATAGCCTTTCATATACCTGTGTTTGTGCTGCGCAGAGTTTA 119578
1790 rThrGluArgPheGlyAspAspValAlaGluIleLysAspSerTyrP 1807
|||||
119577 CTCAGAGCAGAGTTCGGAGATGACGTGTGTAATAATTATCAAGACATCCAAATC 119528
1807 rovalAspLysSerLysLeuAspSerGlnLys..... 1817
|||||
119527 CAGTAGACAAAGAACTGATCCCAACAAAGTGCCGTGATATTAGGC 119478
1817 1817
119477 CTCATGCCATAGGTGCTCCTGCAGTTTGTAATATTCTTATGAGTGT 119428
1818 AlaTyrIleGlnIleThrTyrValGluProTyrPheA 1830
119427 ATGTACTCTTCAGGCCCTACCTCCAGATCACCTACGTCACACCATTTCTCG 119378
1830 sPTThrTyrGluLeuLysAspArgValIthrTyrPheAspArgAspTyrGly 1846
|||||
119377 ACACGATGAGCTAAAGAAAGAGTACCTACTTTGACAAAGACTACAAAC 119328
1847 LeuArgThrPheLeuPheCysThrProPheThrProAspGlyArgAlaNI 1863
|||||
119327 CTGCGTAACCTCATGTACTGTCTCCTCACTGAGCGGGGAGCACA 119278
1863 sGlyGluLeuProGluGlnHisLysArgLysThrIleuLeuSerThrAsn 1880
|||||
119277 CCGCGAACTGAAGAGCAGTACAAACGTAACCACTCTCCACAAAGTGTCT 119228
1880 fAlaPheProTyrIleLysThrArgIleArgValCysHisArgGluGlu 1896
|||||
119227 ACGCTTCCCTTACATCAAGACAGCATCAATGTATATCCACAGAGGAG 119178
1897 Thr..... 1897
119177 GTGTGTTTGAGAGAGCGCAATCTCTAAGTGAATGGATGCGTCCAGC 119128
1898 ValLeuThrProValG 1903
119127 AGGGGTATAAATTTGTTGCTGTGTTTCTCCAGATCAATTTCTCTCCCATAG 119078
1903 LuValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAla 1919
|||||
119077 AGTGGCCATTGAGGACATGCAGAAAGAGACACAGAGTGGCTTGTGCC 119028
1920 ThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnI 1936
|||||
119027 ACAAAACAGAGACCTGACAGACTCCAAAGATGTTGACAGATGTGCTGCAGG 118978
1936 ySerValGlyProThrValAsnGln..... 1944
118977 CTGCGTGGGACCACTGTCAACCAAGTGCCGTGTTACACAGCTGCTCAGA 118928
1944 1944
118927 TGTATATCAATGAAGTCAAAATGCAATGCCCTTATTATAGGGCTT 118878
1944 1944
118877 TAAAGTCTGTCAGAGTGTATTATGTAGGCGCACACATTCCTGCAGAGA 118828
1945 Gly.ProLeuGluValAlaGlnVal 1952

402 CTTCGCACAGAGAGCTCAAGTTCTTACAAACATTAAAGGAGCGCTTA 451
906 LysLeuLeuHisGluLeuAlaLeuGlnTrpValValSe 919
452 CCACGTAAAGACCTTTTCCACAGAGAGCGCTTTCAGAGTGGGTTTG 501
919 rSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPhePheG 936
502 CAGTGCACCGCTTTCGGAAATCAGCTTTCGACAAACCGCTGTTCTTTTG 551
936 InLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGlnArgLeu 952
552 AATTAAATGGTAAGACATGTGTGACACATTATACCTTAAATGATAACTT 601
953 AspThrProArgLysLeuArgPheProGlyArgPheLeuAspPheLeu 969
602 GAGGCTCCAAAGAAAGTGTTCACAGAGCTTTCATGATGACATTGC 651
969 rAlaLeuValGlySerValGlyLeuGluValIleThrArgValHisLys 986
652 ACCTCTTGTACACAGATTGTGTGATGATATAGTTTCACGATTTCAGAA 701
986 spValGluLeuAlaGluHisLeuAsnAlaSerLeuAlaPhePheLeuSer 1002
702 AACACGAAATGGTTGAGAGACTCAATACAGCGCTTGCAATCTTCTCAT 751
1003 AspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuValArgAl 1019
752 GATCTGTGTCTGTTATGACAGAGAGATTGTTTTTACCGTTATTAATC 801
1019 aHisTrpLysGlnValAlaIleThrArgLeuGlnSerProAsnProAla 1036
802 CGCTATAAACAAGGTCTTCAACAGCTTACTACCATTAACCAATCCAGT 851
1036 LeuLeuLeuThrLeuArgMetGluPheThrArgIleLeuCysSerHisGlu 1052
852 TTCTGTGCTCTTGAGCGCTGGATTCTTACGAAATCATCTGCGATGATAG 901
1053 HisTrpValThrLeuAsnLeuProCysCysProLeuSerProProAlaSe 1069
902 CACTATGTATACATTAACTTACCTCGACGCTTACTCCACCGCATC 951
1069 rProSerProSerValSerSerThrThrSerGlnSerSerThrPheSer 1086
952 TCACATCACCTTCTGTTCTTCTGCAACATCTCAGAGTCTGTGATTTCTTA 1001
1086 eArgAlaProAspProLysValThrSerMetPheGluLeuSerGlyPro 1102
1002 CGAATGTACAGACCCAAAGATGTGCAAATATGTTTGAATATTCGCTGCT 1051
1103 PheArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaIle 1119
1052 TTCGCCCAACAGCATTTATTTGGCAGAGCTTGTTTACAGACAGCGGCT 1101
1119 uAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaI 1136
1102 CATTTTAAACCTGTGATCTGAAGAGCTGTTGATTCATTAAGAAAGTCA 1151
1136 LeSerAlaValHisSerLeuLeuCysGluHisAspThrAspProArgLys 1152
1152 TCAAATATGTATACAAATTTACTCTCCAGTCAAGACTCAGACCCGGGGTAC 1201
1153 AlaGluAlaThrValLysAlaArgValAlaGluLeuLeuTyrLeuProLeu 1169
1202 TCTGACCCCTCAGATTAAGCTCGAGTGGCCATGTTGATCTTACCTGTAT 1251
1169 uSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaGluGly 1186
1252 TGCTATTATCATGGAACCTGATCAGCTGTATGATTTTACAGAAATC 1301
1186 rGlyGlnArgSerArgLeuAlaSerMetLeuAspSerSerThrGluGly 1202

1302 ACAATCAAGAGAGAGCAATTTGTTATACCCACGTATGATTAAGAGT 1351
1203 GluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAlaIleAl 1219
1352 GAGAGC.....GGAATATGATTAAGCCAGACCGCTTCCATGGCATTCCG 1395
1219 aGlyGlyProLeu.....AlaProGlySerArgAlaSerIleS 1232
1396 AGGAGCATTCGCTCCCACTAACAAGCGCTGCAGATTTCTCCTCAACG 1445
1232 eArgGlnGlyProProThrIleSerArgAlaGlyLysAlaLeuSerAlaGlu 1248
1446 CA.....ACGAGTGCAGAGGCACACACACTCTTTTACAGACGA 1483
1249 SerSerArgThrLeuLeuAlaCysValLeuTrpValLeuLysAsnThrG 1265
1484 TCAAATCGAAGCGCTTTGATCTGTCTACTTGGGTTCTCAAAAATGCAGA 1533
1265 uProAlaLeuLeuGlnArgTrpAlaThrAspLeuThrLeuProGluLeu 1282
1534 TGAAACAGTTCTACAGAAAGTGTTCACAGATCTCTCAGTCTTCACACTAA 1583
1282 LyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaIlePheGluTyrLys 1298
1584 ACCGCTATTAGATCTGCTTTATCTGTGTGTGCTTGGTATTAATA 1633
1299 GlyLysLysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerIle 1315
1634 GGGAAAAAAGCTTTTGAACGAATGAATAGCTTACCTTTAAGAAATCAAA 1683
1315 uAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrIleGlyAlaAl 1332
1684 AGACATAGAGCAAAAGCTTGAAAGAGCTATCTTGGAGCATGTGGCCA 1733
1332 rGgGlnGluMetValArgArgSerArg.....GluArgSerPro... 1344
1734 GGCAGAAATGCTAGCGCGACGCCGAGACAGCTCCAGAGAACCCCATCT 1783
1345PheGlyAsnProGluAsnValArgTrpArgLysSerValTh 1358
1784 GGAAGTCCCTTGGAAAGTCAAGAAATTTGAGGTGGAGGAAGATATGAC 1833
1358 rHisTrpLysGlnThrSerAspArgValAspLysThrLysAspGluMetG 1375
1834 TCACGTGCGCTCAAAACACAGAGAGCTTGCAATATCAAGAGCGAGATTGG 1883
1375 IuHisGluAlaLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuVal 1391
1884 AACACGAGACAGCTGATTTGATGGAACCTGGCTACAGAGCAACCTTAATC 1933
1392 ValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGluAl 1408
1934 ATTTTATGATTCATTAGAGATTTGTTTCAAGACCGTTTCTGTAAGGAATC 1983
1408 aArgGluSerValLeuGlyAlaValLeuLysValValLeuTyrSerLeuG 1425
1984 CAAAGAGAGCATCTTCTGTGAGAGCTGTAAGAGCTACTACACAGCATGG 2033
1425 LysSerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArg 1441
2034 CCGTATACCAAAAGTGCAGTTTATCTAACACACAGTGTTCCTGACACGAGA 2083
1442 AlaLeuValSerLysPheProGluLeuLeuPheGluGluAlaSerThrGlu 1458
2084 GCCTTGGTTTCAAAAGTTCTCTGTAACCTTATTTGAAGAGAGACAGAGCA 2133
1458 uCysAlaAspLeuCysLeuArgLeuLeuArgHisCysGlySerArgLies 1475
2134 GTGTGCTGATTTATGCTCAGAGCTTCTCCAGACSTGTACAGAGATCG 2183
1475 eThrIleAlaGlnThrHisAlaSerAlaSerLeuTyrIleuLeuMetArgGln 1491
2184 GTACAATAGGCTCACAGCGCAGTGCCTCCCTTACTACTAATAGAGCA 2233

1492 AsnPhgIuIIeGIyHisnPhaIaArgValIlysmetGlnValThrMe 1508
|||||.....
2234 AACTTGGATGGGAATTACTTGGCCAGGGTTAAATGCAGTAACAAT 2283
1508 tSerLeuSerSerLeuValGIThrThrGlnAsnPheserGluGlnHisL 1525
|||||.....
2284 GTCCATATCCCTCTTGGGGCACATCTCAGAAATTTTAAAGAAATTCCT 2333
1525 euATArgSerLeuIlysthrIleuThrThyTAlaGluGluAspMetGly 1541
|||||.....
2334 TAAACGCTTCTAAAGACTATATTGCATATGCTGAAGAAAGACTGGA 2383
1542 LeuArgspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLe 1558
|||||.....
2384 TTGAGGGAAACACATTTCCGATCAGTCAGTCAGGATCTGGTTTCATCT 2433
1558 uHISmetIleuThrAspThrValIlysmetGluHisGluGluAsp 1575
|||||.....
2434 CCATATGATTTCTTCTGATCTGAAATAAGAAACACAGAGAGATC 2483
1575 roGluMetLeuIAspLeuMetThyArgIleAlaArgIlyThrGlnGly 1591
|||||.....
2484 CTGAATGTTGATGATCTAATGTACAGAAATGCCAAGGGTTACACAGCC 2533
1592 SerProAspLeuAArgLeuThrThrLeuGlnAsnMetAlaGlyLysHisAl 1608
|||||.....
2534 TCTCCAGATCGCATGACCTGGTTCAGAAACATGGCAGCAGCAAGCTC 2583
1608 aGluLeuGlyAsnHisAlaGluAlaIleGlnCysMetValHisAlaIle 1625
|||||.....
2584 AGAAGCAAGCAATCATGCTGAAGCTGCACAGCTCTAGTCCACCTCACAG 2633
1625 IAluValAlaIleGlyLeuAlaLeuLeuGluAspGlnArgHisLeuPro 1641
|||||.....
2634 CACTTGTTCGTAATATTGAGCATGCTGAGAGACCGAAATATCTTCCT 2683
1642 ValGlyCysValSerPheGlnAsnIleSerSerPheValIleuGluGly 1658
|||||.....
2684 GTGGAGTGTGAACATTTTCAGAAATTTTCATGATGTTTGAAGAAATC 2733
1658 rAlaIleSerAspAspIleLeuSerProAspGluGluGlyPheCysSerG 1675
|||||.....
2734 TGCAGTCTCAGATGATGTGATCTCCAGATGAAGAAAGTATCGCTCTG 2783
1675 IyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaIle 1691
|||||.....
2784 GAAATATCTTACTGAGCTCAGACTTGTGGGATTTACTGAAACAAGCAGCT 2833
1692 GLyThrPheThrMetGlyGlyLeuThrGluAlaValAsnGluValThrLy 1708
|||:|.....
2834 GCTTCCTTCTCTAAGCTGGCATGTATGAAGCAAGTATGAAAGTTTACAA 2883
1708 sAsnLeuIleProTleuGluAlaHisAspThyLysLysLeuAlaIle 1725
|||||.....
2884 AGTACTATTCTCTATCATGAGCTAATCGGAGATGCAAGAAACTATCCA 2933
1725 IAluValHisGlyLysLeuGlnGluAlaPheThrLysIleMetHisGlnSer 1741
|||||.....
2934 CAATTCATGTGAACCTTCAAGAACATTTCAAGCAAAATTTGTTCAACAAGT 2983
1742 SerGlyThrArgValPheGlyThrThrPheArgValGlyPheThrGly 1758
|||:|.....
2984 ACTGGCTGGAGCGATGTTGGACCTATTTTCGTGTGGTTTATATGG 3033
1758 yAlaHisPheGlyAspLeuAspGluGlnGluPheValThyLysGluPro 1775
|||:|.....
3034 AACCAAGTGGGGGATTTGGATGAACAAGATTTGTTACAAAGAGCGCTG 3083
1775 erIleThrLysLeuAlaGluIleSerHisArgLeuGluGluPheThr 1791
|||:|.....
3084 CAATTAACCAAACTTGCAGAGATATCTCAGAAATTTGAGGAGATTTTACGGA 3133

1792 GluArgPheGlyAspAspValAlaGluIleIleLysAspSerThrProVal 1808
|||||.....
3134 GAAAGATTTGGAGAGAGATGTGTTGAATTAACAAAGACTTAATCCCTG 3183
1808 IAspLysSerLysLeuAspSerGlnLysAlaThrIleGlnIleThrTy 1825
|||||.....
3184 AGACAAGTGTAAATTAGATCTAACAAAGCATATATTCGATTACCTATG 3233
1825 aLGIuProThrPheAspThrThyGluLeuLysAspArgValThrThrPhe 1841
|||||.....
3234 TGGAGCCATACCTTTGCACATATGAGATGACAGACAGAAATCCATTTTC 3283
1842 AspArgAsnThyGlyLeuArgThrPheLeuPheCysThrProPheThrPr 1858
|||:|.....
3284 GACAAAAATTAACATCTCTGATGATCATGTACTGATACCTTTACTTT 3333
1858 oAspGlyArgAlaHisGlyGluLeuProGluGlnHisLysArgLysThrL 1875
|||||.....
3334 AGATGGCCCGCCCATGGGGAACCTTCATGACAAATTCAAAGAGACCA 3383
1875 euLeuSerThrAspHisAlaPheProThrIleLysThrArgIleArgVal 1891
|||:|.....
3384 TTCTGACTACGTCTCATGCTTCTTCTTATATTAACAAAGGCTCAATGTC 3433
1892 CysHisArgGluGluThrValLeuThrProValGluValAlaIleGluAs 1908
|||:|.....
3434 ACTCATTAAGAAAGATCATCTTAACACCAATTAAGTTGCTATTGAGGA 3483
1908 pMetGlnLysLysThrArgLLeuAlaPheAlaThrGluGlnAspProp 1925
|||||.....
3484 CATGCAGAAAGAACACAGAGATGGCATTTGCCAACACATCCAGATCCCG 3533
1925 roAspAlaLysMetLeuGlnMetValIleuGlnGlySerValGlyProThr 1941
|||:|.....
3534 CAGACCCCAAAATGCTTCAGATGTGCTCAGAGATCTGTAGCACCACA 3583
1942 ValAsnGlnGlyProLeuGluValAlaGlnValPheLeuAlaGluIlePr 1958
|||||.....
3584 GTGAATCAGAGGGCCTTGGAACTGGCCAGGTTTCTGTGGAATATCC 3633
1958 oGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeuCysPheL 1975
|||:|.....
3634 TATGACCCAAAGCTCTTACAGATCATATTAATTAAGTGCACCTCTTTA 3683
1975 yAspPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLeu 1991
|||||.....
3684 AAGATTTTACTTAAAGGTGTAAGATGCTTAAGAAAAAATTAAGAGCTTA 3733
1992 ILeGlyProAspGlnLysGluThrHisArgGluLeuGluArgAspThrCy 2008
|||||.....
3734 ATTTGGCCGGATCAAAAGAGATATCAAAAGGAAGTGGAGAAACTATCA 3783
2008 sArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGlnL 2025
|||||.....
3784 TGCCTTAAGAGGCCCTTACAGCCACTGATCAACAGAAAGACCTCAGT 3833
2025 euMetAlaProThrProProGlyLeu.....ArgAsnSerLeuAsnArg 2039
|||:|.....
3834 TATACAGGACAGTATTGCTGTACCTGTCCACAGAGATTCCTTCAGTGA 3883
2040 AlaSerPheArgLysAlaAspLeu 2047
|||:|.....
3884 ATGAGCCTTCGCAAAATGGAATCTC 3907

seq_name: gb_hlg:AC017375
seq_documentation_block: 73515 bp DNA linear HTG 09-DEC-1999
LOCUS AC017375
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces.
ACCESSION AC017375
VERSION AC017375.1 GI:6553611
KEYWORDS HTG; HTGS_PHASE2.


```

416 .....AspSerGluGluArgProAl 424
      ::::::::::::::::::::
41730 GGCAGAGCTTGGCGACAGCGCCAACTCCAACTGATGATCGAATAAC 41681
      ::::::::::::::::::::
424 aATpThr.....AspArgArgArgGlyProGlnAsp..... 435
      ::::::::::::::::::::
41680 TTCACACAGTAGCTTGATCATCACTAAGCGGAAAGCTAACGATGAGCG 41631
      ::::::::::::::::::::
436 .....ArgAla 437
      ::::::::::::::::::::
41630 GAACACTACAGCAGCTGGTTCGCTTGAACGAGGAAAGCAACGCGATCC 41581
      ::::::::::::::::::::
438 SerSerGlyAspAsp...AlaCysSerPheSerGlyPheArgProAlaThr 453
      ::::::::::::::::::::
41580 TGGTCCCGCAGATGATTTGCAAAAGCTTTGAGAAACCTTTCGCTCATAC 41531
      ::::::::::::::::::::
453 rLeuThrValThrAspPhePheGlnGluAlaGluArgLeuSerAspG 470
      ::::::::::::::::::::
41530 TATACAGTACCAAGCTTCTTTAAACAGAGCGCATTAATAAGAAAGACG 41481
      ::::::::::::::::::::
470 LuAspLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeu 486
      ::::::::::::::::::::
41480 AGGATTTGTACAAATAATTTACAGAGTAAACGACCGAGCTCAGTTATG 41431
      ::::::::::::::::::::
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
      ::::::::::::::::::::
41430 AAGAGTACAAATGCAATACCTGGTCTATTAATTAAGAAATATACCTTGG 41381
      ::::::::::::::::::::
503 aProGluAsnProHisPheCysLeuSerProGluLeuHisIleLysP 520
      ::::::::::::::::::::
41380 CGTCGAGGAAAGCCAGATGCAATTAACGCCAGAGTTAGCAACCATTAAC 41331
      ::::::::::::::::::::
520 rOTyrProAspProArgGlyArgProThrLysGluLeuLeuGluPhePro 536
      ::::::::::::::::::::
41330 CTCGAAAGTCAGATTAATGTCGCCCGCAGTTAAGGAAATCTTAGAATTTCCG 41281
      ::::::::::::::::::::
537 AlaArgGluValThrAlaProHisThrSerTyrArgAsnLeuLeuLys 553
      ::::::::::::::::::::
41280 CAATCGCGGATTTACATCCACATTTAGCTATCGCAATCTCTATTTGTG 41231
      ::::::::::::::::::::
553 lTyrProHisSerLeuAsnPheSerSerArgGlnGlySerValArgAsnL 570
      ::::::::::::::::::::
41230 GTCCACCAAGAGCTAAATTTTCTCGCGTGCAGGTTCTGCTCGCAACA 41181
      ::::::::::::::::::::
570 euAlaValArgValGlnTyrMetThrGlyLeuAspProSerGlnAlaLeu 586
      ::::::::::::::::::::
41180 TTGCCGTTCCGTTCAACTAATGCTGAGAAAGCCAAAGATCGCGTC 41131
      ::::::::::::::::::::
587 ProValIlePheGlyLysSerSerGlyPheThrArgGluAlaThr 603
      ::::::::::::::::::::
41130 AATGCCATCTACGCAAGTGTCTGTCCCAATTTTCTAAGCAACATTT 41081
      ::::::::::::::::::::
603 eThrProValValTyrHisAsnLysSerProGluPheTyrGluGluPheT 620
      ::::::::::::::::::::
41080 TACAGCGCTCAATTAATCAACAATGCCCATCATCTTACGATGAATAA 41031
      ::::::::::::::::::::
620 yLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuPheThr 636
      ::::::::::::::::::::
41030 AAATGGCGCTGCTGATCAATAAAGCAGCATCATTTATGTGTACCC 40981
      ::::::::::::::::::::
637 PheTyrHisValSerCysGlnProArgPro.....GlyThrAlaLe 650
      ::::::::::::::::::::
40980 AATTTACCATGTTCATGTCAAAAGAAACCCAGGACTTGCAGCCGCGAGT 40931
      ::::::::::::::::::::
650 uGluThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg 667
      ::::::::::::::::::::
40930 AGAGACTCCATAGCGGTACACATGTTGCCCTTACTGGAAGATGGAAC 40881
      ::::::::::::::::::::
667 euArgThrGlyProPheCysLeuProValSerValAspGlnProProP 683
      ::::::::::::::::::::
40880 TTAAAGTTGGGAAATTAATCTTCGTGTATGTAAGATCAGCCGCGAA 40831
      ::::::::::::::::::::
684 SerTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTrpVal 700

```

```

      ::::::::::::::::::::
40830 AATTACTCATTTATPCCACGAAATGTCACCTGCTCGTAATTAATGCT 40781
      ::::::::::::::::::::
700 lAspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerV 717
      ::::::::::::::::::::
40780 GGCACATACACAGACCGCTGTTCTTATTAATGATGAGCAATGCCCA 40731
      ::::::::::::::::::::
717 aHisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisVal 733
      ::::::::::::::::::::
40730 TTTCATCTTGTGATCTTCTTCGATCGATCTTCTTAATTTGCGAATAT 40681
      ::::::::::::::::::::
734 LeuGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerG 750
      ::::::::::::::::::::
40680 CTAGCACTCGAAATATACCT.....TCCCATATTGGCGA 40646
      ::::::::::::::::::::
750 uGlyAsnValGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeu 767
      ::::::::::::::::::::
40645 GAATTAATATACAAACAGATTAAGAAATGCTTCTGATTCGAATATG 40596
      ::::::::::::::::::::
767 lAserProGluProLeuValAlaPheSerHisValLeuAspLysLeu 783
      ::::::::::::::::::::
40595 CAATTCGTGAGCCATTAAGTGGCATCTACCATTAAGCTTGGATTAAGCTT 40546
      ::::::::::::::::::::
784 ValArgLeuValIleArgProProIleIleSerGlyGlnIleValAsnLe 800
      ::::::::::::::::::::
40545 ATCGAATTTGTTAGTTGTACCCACAAAGTTGTGACACAGCAATGTCTCT 40496
      ::::::::::::::::::::
800 uGlyArgGlyAlaPheGluAlaMetAlaHisValSerLeuValHisA 817
      ::::::::::::::::::::
40495 GGGATCTACGGTTTTCGAAGTTCTTGTGTGATCACTTGTATTACAA 40446
      ::::::::::::::::::::
817 rGSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAla 833
      ::::::::::::::::::::
40445 TACTTAACGAGCATCAATATATGACCAATAGCGGCGCAAAAGTTGCTATCA 40396
      ::::::::::::::::::::
834 AlaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPr 850
      ::::::::::::::::::::
40395 ACATATGTGCACCTTTCAATGTAAATTTCCGCA..... 40364
      ::::::::::::::::::::
850 cAspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArg... 865
      ::::::::::::::::::::
40363 .....CCCTTTCAAAAGCAACAGCGACCTACATCGAGTCGC 40327
      ::::::::::::::::::::
865 ..... 865
      ::::::::::::::::::::
40326 AGCACACAGAAAGACTTGCATTAAGCCGAAATGTAACCATATATGATTA 40277
      ::::::::::::::::::::
866 .....GlySerGlyArgProAlaSerLeuTyr 874
      ::::::::::::::::::::
40276 TGTTTAGCAAGTGGCCGAGCTGTGATCGCAAGG...TAAATATACATA 40230
      ::::::::::::::::::::
874 rLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAlaVala 891
      ::::::::::::::::::::
40222 TTTTGTAAAGCTTAAATATATACCTTTTCCCTTCAAGCTGTGCATTA 40181
      ::::::::::::::::::::
891 lAProGlySerValAspAspGluValSerArgIleLeuAlaSerLysLeu 907
      ::::::::::::::::::::
40180 ....GATGTCTTCACCTCGATGCTGCACGAGTGCAGATGTCCAAAGTCCGGCTG 40136
      ::::::::::::::::::::
908 LeuHisGluGluLeuAlaLeuGlnTrpValValSerSerSerAlaValArg 924
      ::::::::::::::::::::
40136 CTTCACGAAAGATTAAGCTTGCATTTGGGTGTGTCAMATGGGAAAGCGAGC 40086
      ::::::::::::::::::::
924 gGluAlaIleLeuGlnHisAlaTrpPhePhePheGlnLeuMetValLys 941
      ::::::::::::::::::::
40086 TGACCTTAGCAGATGCAAAATGCTGTTTATTTAGCTTATTTGTGCATAT 40036
      ::::::::::::::::::::
941 eMetAlaLeuHisLeuLeuGluGlyGlnArgLeuAspThrProArgLys 957
      ::::::::::::::::::::
40036 CTATGATTTGAACATTTGATTTGTTCAAAATACTTTAAAGTGGCCCGGAAA 39986
      ::::::::::::::::::::
958 LeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySe 974
      ::::::::::::::::::::

```


1526 rgArgSerLeuYsrhrlleuThrTyralaGluGluaspMetGlyLeu 1542
 |||||.....
 38205 GCCGTCGCTTAACCGCTCGGTTATGCTTAATCGATCCGACCTG 38156
 1543 ArgAspSerThrPheAlaGluGlnValGlnaspLeuMetPheasnLeu1 1559
 :|||:|||||.....
 38155 CAGGATACTTCTTTCTCGAACACATACAGATTGCTTTTATCTGCA 38106
 1559 sMetlleuThrAspThrValLysMetLysGluHisGlnLysProG 1576
 |||||.....
 38105 TATGATCCCTGATGATCTGTTAAATGAGAGATATCGAGAACCCAG 38056
 1576 lMetleuLeaspLeuMetTyralGlyAlaArgGlyTyrgLysSer 1592
 |||||.....
 38055 AATGTGCTTACCTCATGATGCTATTCGCAAGGATACCAAAATAC 38006
 1593 ProAspLeuArgLeuThrTrpLeuGlnaspMetAlaGlyLysHisAlaG 1609
 |||||.....
 38005 CCGATCTACGACTGACTTGTTGAAAAATATGCTAAAAAACCCGCA 37956
 1609 uLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaL 1626
 | :|||:.....
 37955 GCGAGCAATTCACAGGAGACGCAATGTTATGATACATGCTGCTCT 37906
 1626 euValAlaGluTyrlleuAlaLeuLeuGluaspGlnArgHisLeuProVal 1642
 |||||.....
 37905 TGTTTCTCAATATCTTAGCATGTGGAGTCCAAAAACATTTGCTGCTT 37856
 1643 GlYCysValSerPheGlnAsnLleSerSerAsnValLeuGluGluSerAl 1659
 |||||.....
 37855 GAGAGCTGTAAGTTTTCACGAAATTTCTCCACACACATTAAGAGTCCGC 37806
 1659 aLleSerAspAspLleLeuSerProAspGluGluGlyPheCysSerGlyL 1676
 |||||.....
 37805 CGATTCGATGATGTCCTAAGTCCCGGCAAGATGCTATCTGCCAGGAA 37756
 1676 yshIsphethrGluLeuGlyLeuValGlyLeuGluGluAlaAlaL 1692
 :|||:|||||.....
 37755 ATCATTTCACTGAAACTGGGTTGAAGGCTTCTGGAAGAGCCCTCCAT 37706
 1693 TyrPheThrMetGlyGlyLeuTyrgLualaValaAsnGluValTyrlYsAs 1709
 ||| :|||:.....
 37705 TCCTTTCAGTTGCTGCGATGTATGAGCAATGAACGAAGTGTACAAAT 37656
 1709 nLeuLleProLleLeuGluAlaHisArgAspTyrlYsLysLeuAlaLay 1726
 |||||.....
 37655 TCTAATACCCATATGCGAGGCTAACAGAGATTTTCAAAAAGCTAAAGCAAG 37606
 1726 aHisGlyLysLeuGlnGluAlaPheThrLysLle..... 1737
 |||||.....
 37605 TTCATGCAAAATTGCGAGAGGCAATTAATCGAATATCCCAACTACAGT 37556
 1738MetHisGlnSerSerG 1743
 |||||.....
 37555 ACAATATGTGTAAATTTTACCAACGAAATATATACATATTATTAACA 37506
 1743 yLTPGluAcGValPheGlyThrTyrlPheArgValGlyPheTyrgLualH 1760
 :|||:|||||.....
 37505 GGGTAGAGAGGTTTGGAAACATCTTGTGCTTGTGCTTATGCGGAA 37456
 1760 lAspHeGlyAspLeuaspGluGlnGluPheValTyrlYsGluProSerL 1776
 |||||.....
 37455 AATTTGGGCTTGTGATCAGCAGCAATTCATTTATTAAGACCAACATG 37406
 1777 ThrLysLeuAlaGluLleSerHisArgLeuGlu..... 1787
 |||||.....
 37405 AGGAGTGTGCCGAAATATTAGTGGCTTCAGGTATATATGCAAAATG 37356
 1788Glu..PheTyr 1790
 ||| |||||
 37355 GAAAAATAGAACATATCAATTTTGTTCACATACGTTAGAACCTTTAC 37306
 1791 ThrGluArgPheGlyAspAspValValGluLleLleLysAspSerTyrPr 1807

|||||.....
 37305 ACTGACCACTTCGACCGGACTCTGTGATATCATTAAGATTCCAAATAC 37256
 1807 oValAspLysSerLysLeuaspSerGlnLysAlaTyrlleGlnLleThrT 1824
 ||||| :|||:.....
 37255 CGTTATATATATAGCTTGTGATCCCGATTAAGGCTTACATTCAAATTTACT 37206
 1824 yValGluProThrPheAspThrTyrgLuleuLysAspArgValThrTyr 1840
 |||||.....
 37205 ATGTTGAACCTTACTTGAACATATGAAATGCGTCATCGTGAACATAC 37156
 1841 PheAspArgAsnTyrgLyeu..... 1847
 |||||.....
 37155 TTTGAGCGGAATTCATATATAGATGATGATGAATTAACAGATAATTTA 37106
 1848ArgThrP 1850
 |||||
 37105 AATCGAAATTTAAATTTATATTTTAACTTTTGGTTTATGCAAGCT 37056
 1850 heLeuPheCys..ThrProPheThrProAspGlyArgAlaHisGlyL 1866
 || :|||:|||||.....
 37055 TTTATATATGCGACGCTTTTACTTAAATGAAAGCAGCATGCGCAACT 37006
 1866 uProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1883
 | |||||.....
 37005 AATGAGCAGTGTAAACGAAACGATTTTGACACGCGGATACCTTTC 36956
 1883 rOTyrlleLysThrArgLleArgValCysHisArgGluGluThrValLeu 1899
 |||||.....
 36955 CATAGCTAAAACTCGCATTTATGCTAATTTAGCCGACACCAAAATGCTCTG 36906
 1900 ThrProValGluValAlaLleGlu..... 1907
 |||||.....
 36905 GAACCGATTGAAGTGCAATTTAAGTAAGTATGTTGTAATAATCCCAAA 36856
 1908AspMetGlnLysLysThrArg 1915
 |||||.....
 36855 GTTAATAAATTAACGTTATGCTAGATATTCAAAAAACATTTG 36806
 1915 luleuAlaPheAlaThrGluGlnaspProAspAlaLysMetLeuGln 1931
 ||||| |||||.....
 36805 AATTTGCACCTGCCAGATTCAGAGCAGCAGACCCCAAAATTTGCA 36756
 1932 MetValLeuGlnGlySerValGlyProThrValaAsnGlnGlyProLeuG 1948
 |||||.....
 36755 ATGCTACTGACAGGATGTATTTGGAACCACTGTTAACCAAGACCGATG 36706
 1948 uValAlaGlnValPheLeuAlaGluLleProGluAspProLysLeu...P 1964
 |||||.....
 36705 AATGGGAGAGTGTCTTCCATTTATCCGACGAGCAACTGTACCGA 36656
 1964 heArgHisAsnLysLeuArgLeuGlyPheLysAspPheCysLysLys 1980
 :|||:|||||.....
 36655 CAAACACCAAAACAACTTCGTTGCTGCTTGGGAGTTTCAAGGCT 36606
 1981 CysGluAspAlaLeuArgLysAsnLysAlaLeuLleGlyProAspGlnL 1997
 ||| |||||.....
 36605 TGTGCGATAGCTTTGAAAAGATCGCAATCTAATATCTTACAGACAAA 36556
 1997 sGluTyrlHisArgGluLeuGluArgAsnTyrlCysArgLeuArgGluAlaL 2014
 |||||.....
 36555 AGATTACCAACGAGATCGTAACGTAACACAGATCGTTTCAATTGAACAT 36506
 2014 euGlnProLeuLeuThr 2019
 || |||||.....
 36505 TGACTCCCTTTATTTACT 36489
 seq_name: gb_in:AC008318
 seq_documentation_block: 166626 bp DNA linear INV 05-APR-2001
 LOCUS AC008318
 DEFINITION Drosophila melanogaster, chromosome 2L, region 21A-21A, BAC clone
 BACR05J20, complete sequence.

ACCESSION AC008318
 VERSION AC008318.7 GI:13549312
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 166626)
 REFERENCE
 AUTHORS Celniker, S.E., Adams, M.D., Krommiller, B., Tyler, D., Wan, K.H.,
 Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Brandon, R.C.,
 Rogers, Y., An, H., Baldwin, D., Banzon, J., Beeson, K.Y., Busam, D.A.,
 Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M.,
 Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Drenek, D., Farfan, D.,
 Ferreira, S., Frise, E., Galle, R.F., Gary, N.S., George, R.A.,
 Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J.,
 Ibegwam, C., Jalali, M., Kruse, D., Li, P., Maltel, B., Moshrefi, A.,
 McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J.,
 Pacle, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B.,
 Phanuenavong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F.,
 Stapleton, M., Strong, R., Svitskas, R., Tector, C., Williams, S.M.,
 Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
 Sequencing of Drosophila chromosome 2L, region 21A-21A
 Unpublished
 2 (bases 1 to 166626)
 TITLE
 JOURNAL Celniker, S.E., Agbayani, A., Arcalata, T.T., Baxter, E., Blazey, R.G.,
 REFERENCE Butenoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 AUTHORS Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
 Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Shit, E.,
 Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
 Rubin, G.M.
 Direct Submission
 TITLE Submitted (02-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
 JOURNAL Laboratory, MS 64-121, Berkeley, CA 94720, USA
 COMMENT On Apr 5, 2001 this sequence version replaced g1.6728945.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 This sequence was assembled using end sequences from a whole genome
 shotgun and from subclones of this BAC and its neighboring clones.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdgpre@fruitfly.berkeley.edu.
 FEATURES
 source
 location/Qualifiers
 1. 166626
 /organism="Drosophila melanogaster"
 /strain="y: cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="2L"
 /map="21A-21A"
 /clone="BACR09J20 (DB41)"
 /clone_1lb="RPC1-98 (Rosewell Park Cancer Institute
 Drosophila melanogaster BAC library, partial EcoRI in
 PBAC3.6)"
 BASE COUNT 49737 a 32951 c 34017 g 49921 t
 ORIGIN
 Alignment_scores: Quality: 4423.00 Length: 2260
 Ratio: 2.881 Gaps: 45
 Percent Similarity: 67.920 Percent Identity: 43.761
 alignment_block:
 us-09-736-968a-2 x AC008318 ..
 Align seg 1/1 to: AC008318 from: 1 to: 166626

```

13 11aasnargthrvalaalaagluvalarglysglnvalserargsluar 29
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85623 ATCCCTAGCGACGTCGGCTGGCGATGTCGAAAAAATCTTCTGGCCA 85672
29 gserglyserprohisserargargycyserserleuglyvalp 46
: : : : : : : : : : : : : : : : : : : : : : : : :
85673 TCTATCAAGCAATGAGCAACCATCATGTGTGCTTCGTAAGAGCAC 85722
46 roleuthrgluval.val..... 51
|||
85723 TATGTGTTAAAAATATACAAATTTTCTGTAGACATTCATTTTCAAA 85772
52 .....GluProLeuaspPheGluaspValLeuLeuSerArgp 64
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85773 GATCTCCCGACAGCCCTGTATTCGAAGAATTTCTATCCACATTA 85822
64 roProaspAlaGluProGluProLeuArgaspLeuValGluPheProAla 80
: : : : : : : : : : : : : : : : : : : : : : : : :
85823 TGAACATCATCAACCCGAGTCCCTGAAACATATATTGACCTTCCCAA 85872
81 AspaspleuGluLeuLeuGluProArgGluCysArgThrGluPur 97
||||| : : : : : : : : : : : : : : : : : : : : : :
85873 GGGGATGTAAGTCTTAATATCATTCGCCCAAAATCGAATCTGTATCA 85922
97 OGlyIleProLysaspGluLys.leuaspAla..... 107
: : : : : : : : : : : : : : : : : : : : : : : : :
85923 CGTTGTTCCCAATGAGAAATTTGTATGTACACCTAATTAACACAGC 85972
108 .....GlnValArgAla 111
TACATAGCTTACAGTCTCTTTAGTGTGATTTGCCCATACAGTCACAGCA
85973 TACATAGCTTACAGTCTCTTTAGTGTGATTTGCCCATACAGTCACAGCA 86022
111 aalaValGluMetYrIleGluaspTrpValIleValHisArgArgYrG 128
: : : : : : : : : : : : : : : : : : : : : : : : :
86023 ATGTGTTAACTGATATACGGCCCATGGAAGTGTGAATATGACACAGC 86072
128 InTYrLeuSerAlaAla..... 133
: : : : : : : : : : : : : : : : : : : : : : : : :
86073 GGCACTTTCACAGTCTGTGATATCCGCGAGCCTATACAGCCGGAGAC 86122
134 TyrSerProValThrThrAspThrGlnArgGluArgGlnLysGlyLeuP 150
||||| : : : : : : : : : : : : : : : : : : : : : :
86123 ATAGTCTCCCTGGTACACAGCAAGATTCGAAATCGCAAGATTTT... 86170
150 cArgGlnValPheGluGlnaspAlaSerGlyaspGluArgSerGlyProg 167
: : : : : : : : : : : : : : : : : : : : : : : : :
86171 ...TCGTCATTTGAGAGGCGGTTGCTATTAAGAGAGAGCTGCACCTC 86216
167 LuaspSerAsnAspSerArgArgGlySerGlySerProGluAsp..... 181
: : : : : : : : : : : : : : : : : : : : : : : : :
86217 CAAAGTCTCGACAAATCCATTGCCAAGTTTGCTTCAGTAAGCTTTGTACG 86266
182 .....ThrProArgSerSerGlyAlaSerSerIlePheaspLeuAr 195
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86267 GACACTTTCAGCCCGCGGTGTTCTTGCGGTACG.....TTGATCTTCG 86310
195 gasLeuAlaalaaspSerLeuLeuProSerLeuLeuGluGlnArgAlaIar 212
: : : : : : : : : : : : : : : : : : : : : : : : :
86311 ACGTTCGGTATATGACCCCATTAATCCCAATTTGCTTACATATATACCTC 86360
212 roGluaspValAspArgAsnGluThrLeuArgArgGlnHisArgPro 228
||||| : : : : : : : : : : : : : : : : : : : : : :
86361 CGGACATATTTGATCAATCCAAATGACAGCAGCCGGGACAGAAATCGACAA 86410
229 ProAlaLeuLeuThrLeuTyProAlaProAspGluaspGluAlaValGI 245
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86411 ATGGCTTTTCTCAGTATCCGGAATCTGAAGAGGAGAAACATTTGA 86460
245 uArgCysSerArgProGluProProArgGluHisPheGlyGlnArgIleL 262
||||| ||| ||| ||||| ||||| ||||| ||||| ||||| |||||
86461 ACGCCGCTGCTCGCAGAGATTCAGTTGAACATATGGGTATCGCATTC 86510
262 euValLysCysLeuSerLeuLysPheGluIleGluIleGluProIlePhe 278

```



```

834 AlaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuPr 850
      ::::::::::::::::::::
88322 ACATATGTGCACTTCAATGTAAATTCGCA..... 88353
850 oAspGlyAlaProPValThrValGlnAlaAlaThrLeuAlaArg.... 865
      ::::::::::::::::::::
88354 .....CCCTTTCAAAGCAAAACAGGACTAACATGCACTGCC 88390
865 ..... 865
88391 AGCACAACAGAACCTTGCAATTAGCGAATCGTACACCATATATGATAA 88440
866 .....GlySerGlyArgProAlaSerLeu 874
88441 TGTTTAGCAAGTGGCCGAAGCTTGATGCGCAAG...TAAATATCAT 88487
874 rLeuAlaArgSerLysSerLysSerSerSerAsnProAspLeuAlaVal 891
      ::::::::::::::::::::
88488 TTTTGTAAAGCTTTAAATATACCTTTTCTTCAGAGCTGTCATA.. 88536
891 lAProGlySerValAspAspGluValSerArgLLeuAlaSerLysLeu 907
      ::::::::::::::::::::
88537 .....GATGCTCTCACTCGATGCGTGCAGCGGATGTCACAGTGGCGTG 88581
908 LeuHisGluGluLeuAlaLeuGlnTrpValValSerSerSerAlaValAr 924
      ::::::::::::::::::::
88582 CTCACACAATAATTAGCTTGATGCGTGTGTCAGAAAGTGAAGGACAG 88631
924 gGluAlaIleLeuGlnHisAlaTrpPhePheGlnLeuMetValLys 941
      ::::::::::::::::::::
88632 TGACTTAGCCATGTCAAAATCGTGTTTATTTGAGCTTATGTCAAAAT 88681
941 eMetAlaLeuHisLeuLeuGlnGlnArgLeuAspThrProArgLys 957
      ::::::::::::::::::::
88682 CTATGATTGAACATTTTCATTTGCTTCAATACCTTAAATGTCGCCGCAA 88731
958 LeuArgPheProGlyArgPheLeuAspPheThrAlaLeuValGlySe 974
      ::::::::::::::::::::
88732 CATGATTTCCACATCAGTTAATGATGATATCCACCCTGCGACATT 88781
974 rValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaG 991
      ::::::::::::::::::::
88782 AGTTACACAACAAAGTGTGCGCTACACACACCAACCTAAATTAGCGC 88831
991 LuHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeu 1007
      ::::::::::::::::::::
88832 AGTCATTAATGCAAGTTTAGGCTTTTATTTATTTGACATCTAACGATC 88881
1008 ValAspArgGlyPheValPheSerLeuValArgAlaHisTyrLysGlnVa 1024
      ::::::::::::::::::::
88882 ATGATCGGGGATTGTTTGGGCTCATCAAAACGTACACAAGGTTCT 88931
1024 lAlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeu 1041
      ::::::::::::::::::::
88932 GATTTCATAAAATGCTCAATACCGGAT.....CTAATGAACTATA 88972
1041 rGmetGluPheThrArgLLeuGlySerHisGlnHisTyrValThrLeu 1057
      ::::::::::::::::::::
88973 AAATAGACTTTTGAATTTGTGTGTAGTCATGACATTTTGTAGCCTTA 89022
1058 AsnLeuPro.....CysCysProLeuSerProProAlaSerP 1070
      ::::::::::::::::::::
89023 AACTTGGCGTTTGTACTTCATACAAATGTAAACAGCGCCATGACGCC 89072
1070 oSerProSerValSerSerThrThrSerGlnSerSerThrPheSerSerG 1087
      ::::::::::::::::::::
89073 CAGCCCAAGCAACAGTCGAGCAACAGCAACCTTTGCGTAAGTAAAA 89122
1087 lAla..... 1088
89123 AAACAATACACTTAAATTTTACTTATTTGTGCTTTTAAAGGAT 89172

```

```

1089 .....ProAspProLysValThrSerMetPheGluLeuSerGlyProPh 1103
      ::::::::::::::::::::
89173 CACTAGACAGAGCCCTTCACGCC.....GACCTTAGTAAAGAAAT 89212
1103 eArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeu 1120
      ::::::::::::::::::::
89213 CCTCAGCAACATTTTCTAGTCGAGCTGTTCTAGGCACTTACACACAG 89262
1120 lAlaLeuGlu...ProGluAlaGluGlyAlaPheLeuLeuHisLysLysAla 1135
      ::::::::::::::::::::
89263 TAATGGAGGTGCCAACCACCCCA.....TTACATGAAAAAGCA 89300
1136 lIleSerAlaValHisSerLeuLeuGlyHisAspThrAspProArgTy 1152
      ::::::::::::::::::::
89301 ATAAGCTATATTCGCAACCTGATGACATCTCAGATTGACAGCGCGGTA 89350
1152 rAlaGluAlaThrValLysAlaArgValAlaGluLeuThrLeuProLeu 1169
      ::::::::::::::::::::
89351 TAGCAAACTGATGCGCTGCAAGGCTGCTTTATATATATACCGCTAC 89400
1169 eSerLleAlaArgAspThrLeuProArgLeuHisAspPheAlaGluGly 1185
      ::::::::::::::::::::
89401 TTTCTATAGTGATGATGATTCATCTCAGTTGCATTCAGACAGCGCTGAG 89450
1186 ProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThrGluG 1202
      ::::::::::::::::::::
89451 CAGGATCGTCTGTCAGCAAAATCGTCACTGATGAACTTCAAGCGGCACA 89500
1202 yGluGlyAspLleAlaGlyThrIleAsnProSerValAlaMetAlaIle 1219
      ::::::::::::::::::::
89501 TCAACCTATTACAGCATACACATTAATCCGGAAGTGCATTTGCGAAT 89550
1219 lAglyGlyProLeuAlaProGlySerArgAlaSerLysSerGlnGlyPro 1235
      ::::::::::::::::::::
89551 CGGGAAGTCGTCCTACTCTACTGATGAATGAACAAAGTAAAAAATAATG 89600
1236 ProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlySerSerArgTh 1252
      ::::::::::::::::::::
89601 CCC.....CTCAGCTCGGAAATACCCGCA 89626
1252 rLeuLeuAlaCysValLeuTrpValLeuLysAsnThrGluProAlaLeu 1269
      ::::::::::::::::::::
89627 TTTATTTGGTGTTTCTGCTGCTGCTTAAAAAATCTGGAGCGCAACGTT 89676
1269 eGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLeu 1285
      ::::::::::::::::::::
89677 TATATCGTGGCTTATGATTTAAGTCCGATAGAGTGCACCAATGCTT 89726
1286 AspLeuLeuTyrLeuGlySerLeuAlaAlaPheGluTrpLysGlyLys.. 1301
      ::::::::::::::::::::
89727 CAAGGTGTTAACGTTTGTGTAACCTTTCGAGTACACAGCAACAAAAA 89776
1302 .....AlaPheGluArgLLeuAsnSerLeuThrPheLysLys.....S 1314
      ::::::::::::::::::::
89777 CGTGGCCACCTTAACAGCAATACACAAAGTTTAGAAACTGGAT 89826
1314 eLeuAspMetLysAlaArgLeuGlnGluAlaIleLeuGlyThrIleGly 1330
      ::::::::::::::::::::
89827 CTACTGATGTAAAGAAAAAGCTGGAAGATGATTAGAGCAACAAATCA 89876
1331 AlaArgGlnGluMetValArgArgSerArgLLeuArgSerProPheGlyAs 1347
      ::::::::::::::::::::
89877 GCACGTTAATGTTAATATGCTGAAGAAAGTCTAAT.....TC 89917
1347 nProGluAsnValArgTrpArgLys...SerValThrHisTyrLysGlnT 1363
      ::::::::::::::::::::
89918 AACAGAAAAATTCGATGAGGAAGAACAGACAGATGCGCATGCGTGCAT 89967
1363 hSerAspArgValAspLysThrLysAspGluMetGlnHisGluAlaLeu 1379
      ::::::::::::::::::::
89968 ATGCGGATGCTGTGGAAAAAGTGAAGCATGAACCTGAATTAAGCATTTC 90017
1380 ValGluGlyAsnLeuAlaThrGluAlaSerLeuValValLeuAspThrLe 1396

```

```
.....
90018 ATGAAAGGCTTTGGCACTGAAGTTGCACTTGTCTGATACCT 90067
1396 ugiuileilevalaIlnThrValMetLeuSerIuaIaArgIuSerVal 1413
90068 TGAATAATATTGTTCATGCTGCT.....GCCAACCTTTATCAAAATCTTC 90111
1413 eugIyAlaValLeuLysValValLeuTyrSerIuSerIuSerIuSer 1429
90112 TTGGAACTGTGCTAAAGGTGCTTTCATTTCTTGTCCGGAATCATCT 90161
1430 AlAleuPheugIlnHisGlyLeuAlaThrIlnArgAlaLeuValSerIu 1446
90162 GTTTTGCCCTGCAAAACTTTGCTTCCCTCCCAACGGCGCTTAATCTTAA 90211
1446 spHeProGluLeuLeuPheGluGluAspThrGluLeuCysAlaAspLeu 1463
90212 ATTTCCAAACTTATTTATTTGATGACAGACCCGATATATGTGCGGATTTGT 90261
1463 ysiLeuArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThr 1479
90262 GTCTAATTTTCTGAAACATTTGGGTCCCTATTTGCCAGGAATGATCA 90311
1480 HisAlaSerAlaSerLeuTyrLeuLeuMetArgGluAsnPheGluIleG 1496
90312 CAGGCAAGCTGCATCACTATATTACTAATGACACAAATTTTCAAAATTGG 90361
1496 y.His..... 1497
90362 AAATGTAAGTTTATTTATTTAGTTTATTTAACTAAATTAATGGT 90411
1498 .....AsnPheAlaArgValLysMetGluValThrMetSe 1509
90412 CACTTATTCGTATAGAACTTTGCCGAGTTAAGATGCAAGTGCAAGTGTC 90461
1509 rLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGlnHisLeuA 1526
90462 TTTAAGCTCCCTAGTTGGAACAGTTGCGGTCTTACTGAGACATCCTTAC 90511
1526 rArgSerLeuLysThrIleLeuThrTyrAlaGluGluAspMetGlyLeu 1542
90512 GCGGTGGGCTTAAACCGTCCTGTTATGCTGAATCCGACTCCGACCTG 90561
1543 ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuH 1559
90562 CAGGATACCTCTTTCTCTGACACAGTACAGATTTCTTTTATATCTGCA 90611
1559 smetiLeuThrAspThrValLysMetLysGlnHisGlnGluAspProG 1576
90612 TATGATCCTGTCAGATACGTGTTAAATGAAGATACAGGAAGACCCAG 90661
1576 IuMetLeuIleAspLeuMetTyrArgIleAlaArgGlyTyrGlnGlySer 1592
90662 AAATGTGCTTGTGACCTCATGAATCGTATTTGCCAAAGGATACCAAATAAC 90711
1593 ProAspLeuArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaG 1609
90712 CCTGATCTAGACCTGACTTGCTGGTGAATAATGTGCTAAACACCGCGCA 90761
1609 uLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAla 1626
90762 GCGAGCAAAATCACACGGAACACCACTGTGTATGTACATGCGCTTCTT 90811
1626 euValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHisLeuProVal 1642
90812 TAGTTTCTGATAATCTTACATGTGAGTACAAAACATTTGCTGCTT 90861
1643 GLyValSerPheGlnAsnIleSerSerAsnValLeuGluGluSerAl 1659
90862 GGAAGCTTAGATTTTCAACAAATTTCTCCCAACACACTAATGAGTCGGC 90911
1659 aIleSerAspIleLeuSerProAspGluGluGlyPheCysSerGlyL 1676
.....
```

```
90912 CGTATCGATGATGTGCTAAGTCCCGGCAAGATGATATCTCCCTAGGA 90961
1676 yShIsPheThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaAlaGly 1692
90962 ATCATTTCACTGAAACTGTGTTGAAGCGCTTGGAAGAGCTCCCAAT 91011
1693 TyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluValTyrLysAs 1709
91012 TCTTTCAAGTTGCTGCGATGATGAAACAATGAACGAAGTGTCAAAT 91061
1709 nLeuIleProIleLeuGlnAlaHisArgAspTyrLysLeuAlaAlaVal 1726
91062 TCTATATCCCATATGCGAGGCTTAACAGAGATTTCAAAAGTAGCAAG 91111
1726 aHisGlyLysLeuGlnGluAlaPheThr..... 1735
91112 TTCATGGCAAAATTGCGAGGAGCATTTAATCGAATATCCCACTACAGTA 91161
1736 .....LysIle.MetHisGlnSerSerg 1743
91162 ACAATATTTGTAAATTTTACCACGCAAAATATATACATATTTATTAAC 91211
1743 IYTrpGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAla 1759
91212 AGGTAAGAGAGTTTGGAAACATACCTTTCGTGTGGCTTATGCGCGGA 91261
1760 HisPheGlyAspLeuAspGluGlnGluPheValTyrLysGluProSerI 1776
91262 AAATTTGGGACTTGATGATCAGCAGCAATTCATTTATAAAGCCACAACT 91311
1776 eThrLysLeuAlaGluIleSerHisArgLeuGlu..... 1787
91312 GACGAGATTGCCCGAATATTTAGTCGGCTTCAGGTATATTTGCCAATT 91361
1788 .....Glu.PheTyr 1790
91362 GGAATAATATAGACTAATCAATTTTGTTCACAACATACGTAGCACTTTA 91411
1790 rThrGluArgPheGlyAspArgValAlaGluIleLeuLysAspSerTyrP 1807
91412 CACTAAGCATTTGGAGCCGCACTGTGCTATATCATTAACATTTCCAAATA 91461
1807 roValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThr 1823
91462 CCGTTGATATTTATAGCTTTGGATCCCGATTAAGCCTTACATTTCAATTACT 91511
1824 TyrValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTy 1840
91512 TATGTTGAACCTACTTGTGAACATATGAATGCGTCATCGTAGACATA 91561
1840 rPheAspArgAsnTyrGlyLeu..... 1847
91562 CTTTGAAGCGGAATTTCAATATATAGTATGATATGAATTAACAGATAATT 91611
1848 .....ArgThr 1849
91612 AAATGCAATTTAAATATATATTTTAACATTTTGGTTATGCAAGAACG 91661
1850 PheLeuPheCys.ThrProPheThrProAspGlyArgAlaHisGlyLut 1866
91662 TTTTATATATGCGACCTTTTACTAATAAATGAAGACACATGCGCGAAC 91711
1866 euProGluGlnHisLysArgGlyThrLeuLeuSerThrAspHisAlaPhe 1882
91712 TAAATGACACAGTAAAGCAAAAGATTTTGACACGCGGCAACATCTT 91761
1883 ProTyrIleLysThrArgIleArgValCysHisArgGluGluThrVal 1899
91762 CCATACGTAATAACTCGCATATGTGTAATTAGCCGACAGCAATATGTCCT 91811
1899 uThrProValGluValAlaIleGlu..... 1907
91812 GGAACCGATTAAGTGAAGTGAAGTAAAGTAAAGTAAATCCCAA 91861
```



```
90460 ....TCGCATTGTAAGAGCGGTTCCGCTTATMAAAGCGAGAGCTGCACCTC 90505
      :::: |||||::: :::: ::::||||
167 luaspSerAsnaspSerArgArgGlySerGlySerProGluasp..... 161
      ::::|||||:::|::: ::::|||||
90506 CAAGTCTTCGCAATCCATTCGACAGTTGGCTTCGTCAATACGCTCTGTACG 90555
      ::::|||||:::|::: |||||
182 .....ThiProArgSerSerGlyAlaSerSerIlePheaspLeuAr 195
      ::::|||||:::|::: |||||
90556 GACACTTTGACCCCGCGCTGCTTGGGCTAGC.....TTTGATCTTCG 90599
      ::::|||||:::|::: |||||
195 gaSnLeuAlaAlaaspSerLeuProSerLeuLeuGluArgAlaAla 212
      :::: :::: ||| :||| :||| :||| :||| :||| :||| :|||
90600 ACGTTCGGTTAATGACCCACTAATACCAATTCCTTGACAAATGATCCTC 90649
      :||| :||| :||| :||| :||| :||| :||| :|||
212 roGluaspValaspArgArgAsnGluThrLeuArgArgGlnHisArgPro 228
      :||| :||| :||| :||| :||| :||| :||| :|||
90650 CGGAGCAATATTGATCAATCCAAATGACGCGCGCAAGATCGCAAA 90699
      :||| :||| :||| :||| :||| :||| :||| :|||
229 ProAlaLeuLeuThrLeuTyProAlaProaspGluaspGluAlaValG 245
      :||| :||| :||| :||| :||| :||| :||| :|||
90700 ATGGCTCTTTTCTCAGCTGATCCGAGATCGAAGCGGAGAAACAATGA 90749
      :||| :||| :||| :||| :||| :||| :||| :|||
245 uArgCysSerArgProGluProProArgGluHisPheGlyGlnArgIle 262
      :||| :||| :||| :||| :||| :||| :||| :|||
90750 ACGCGCGCTGCTCGAGAGATTCCAGTTGAACAATATGGTCATCGCATTC 90799
      :||| :||| :||| :||| :||| :||| :||| :|||
262 euValIlyCysLeuSerLeuIlyPheGluIleGluIleGluProIlePhe 278
      :||| :||| :||| :||| :||| :||| :||| :|||
90800 AGGTTAACTGCTACAGCTTGCGCTGAGCGGAGAGTGGAGCCAAATTTT 90849
      :||| :||| :||| :||| :||| :||| :||| :|||
279 GlyIleLeuAlaLeuTyArgPvalArgGluIlyLysLysIleSerGluAs 295
      :::: :::: ||||| :||| :||| :||| :||| :||| :|||
90850 GCATCGATGGCTATTATGATGCTAAGAACGCGAAATAATTTCTGAAA 90899
      :||| :||| :||| :||| :||| :||| :||| :|||
295 nPheTyPheaspLeuAsnSeraspSerMetLysGlyLeuLeuArgAlaH 312
      :||| :||| :||| :||| :||| :||| :||| :|||
90900 TTTTCTATTTTGACATGAAACGAGTACTCTTAAACGCATGTTGTCAAGTC 90949
      :||| :||| :||| :||| :||| :||| :||| :|||
312 lAsGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSer 328
      :||| :||| :||| :||| :||| :||| :||| :|||
90950 ATGTACAGTGGCGCGACATAGTACCCAGAACCCCTCAGCATATTCGAA 90999
      :||| :||| :||| :||| :||| :||| :||| :|||
329 ValThrTyProSerProaspIlePheLeuValIleLysLeuGluLysVa 345
      :||| :||| :||| :||| :||| :||| :||| :|||
91000 ATTTCATATCGAGTAGTATTTATTTCTGCTGATTCGTCTGAAAAAAGT 91049
      :||| :||| :||| :||| :||| :||| :||| :|||
345 lLeuGlnGlnGlyAspIleSerGluCysGluProTyTyrMetValLeuL 362
      :||| :||| :||| :||| :||| :||| :||| :|||
91050 TTTTG...CAGGCTGACATTAACAATTCCTGTCGACCCGCTAT.....CTCA 91090
      :||| :||| :||| :||| :||| :||| :||| :|||
362 lYsgIlyValaspThrAlaLysAsnLysGluLysLeu...GluLysLeuArg 377
      :||| :||| :||| :||| :||| :||| :||| :|||
91091 AAGAA.....GACAAAGATAAATACCGAGAGAAAGTAAAA 91125
      :||| :||| :||| :||| :||| :||| :||| :|||
378 LeuAlaAlaGluGlnPheCysTyhArgLeuGlyArgTyArgMetProPh 394
      :||| :||| :||| :||| :||| :||| :||| :|||
91126 TCCAAACCTGTGAGCTACTGTGAACGACTAGCGCAATAATACCGAATGCCGT 91175
      :||| :||| :||| :||| :||| :||| :||| :|||
394 eAlaThrThrAlaValHisLeuAlaAsnIleValSer..... 406
      :||| :||| :||| :||| :||| :||| :||| :|||
91176 TGCATGAGACAGAAATCTACCTAACAAATGTTTAAATGGCGAATAACTTTG 91225
      :||| :||| :||| :||| :||| :||| :||| :|||
407 .....SerAlaGlyGlnLeuaspArgaspSer..... 415
      :||| :||| :||| :||| :||| :||| :||| :|||
91226 AAAGCAAAAGACGTGGCGCGCGGAGAACGGGATTCCTTTGGAATGCTTGT 91275
      :||| :||| :||| :||| :||| :||| :||| :|||
416 .....AspSerGluGlyGluArgArgProAl 424
      :||| :||| :||| :||| :||| :||| :||| :|||
91276 GGCAGACAGCTTGGCAGACGCGCCCAAGCTCAATAGCTTAGATGGAATAATC 91325
      :||| :||| :||| :||| :||| :||| :||| :|||
424 aTrpThr.....AspArgArgArgGlyProGluasp..... 435
      :||| :||| :||| :||| :||| :||| :||| :|||
```

```
91326 TTCCAAGTAGCTTTGATCAACTAAGCGAAGAAAGCTAATGATGACG 91375
      :||| :||| :||| :||| :||| :||| :||| :|||
436 .....ArgAla 437
      :||| :||| :||| :||| :||| :||| :||| :|||
91376 GAACACTTACACAGCGTGGTTCGCTTGAACGAGAGAAAGCACCGCATTC 91425
      :||| :||| :||| :||| :||| :||| :||| :|||
438 SerSerGlyAspAsp...AlaCysSerPheSerGlyPheArgProAlaTh 453
      :||| :||| :||| :||| :||| :||| :||| :|||
91426 TGGTCCCGCAGATGATTTTGCAAACGTTGTAGAAAACCTTCGTCCTATAC 91475
      :||| :||| :||| :||| :||| :||| :||| :|||
453 rLeuThrValIThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspG 470
      :||| :||| :||| :||| :||| :||| :||| :|||
91476 TATACAGTACCAACAGCTCTTTTAAACAAGGCGACATTAATGAAGAACG 91525
      :||| :||| :||| :||| :||| :||| :||| :|||
470 luAspLeuPheLysPheLeuAlaaspMetArgArgProSerSerLeuLeu 486
      :||| :||| :||| :||| :||| :||| :||| :|||
91526 AGGATTTGTACAAAATTTTACAGAGTTTAAACGACCGAGCTCAGTTATG 91575
      :||| :||| :||| :||| :||| :||| :||| :|||
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
      :||| :||| :||| :||| :||| :||| :||| :|||
91576 AAGAGTACAAATGCAATACCTGCTTCTATTAATAGAAATATACCTTG 91625
      :||| :||| :||| :||| :||| :||| :||| :|||
503 aProGluAsnPProHisPheCysLeuSerProGluLeuLeuHisIleLysP 520
      :||| :||| :||| :||| :||| :||| :||| :|||
91626 CGTCGAGGAGCAAGCAAGATGCATTAACGCCAGATGACCTACCTAATACC 91675
      :||| :||| :||| :||| :||| :||| :||| :|||
520 roTyPProaspProArgGlyArgProThrLysGluIleLeuGluPhePro 536
      :||| :||| :||| :||| :||| :||| :||| :|||
91676 CTCMAAGTCCACATATATGCGCCCGCATTAAGAAATCTTAGAATTTGCG 91725
      :||| :||| :||| :||| :||| :||| :||| :|||
537 AlaArgGluValTyArgAlaProHisThrSerTyArgAsnLeuLeuTyVa 553
      :||| :||| :||| :||| :||| :||| :||| :|||
91726 CAATGGCGCATTTTACATCAATCAATATAGCTATCCAAATCCCTATTTGT 91775
      :||| :||| :||| :||| :||| :||| :||| :|||
553 lTyPProHisSerLeuAsnPheSerSerArgGlnGlySerValArgAsnL 570
      :||| :||| :||| :||| :||| :||| :||| :|||
91776 GTCACCAAGAGGCTAAATTTTCTGCGTGCAGCTCTGTGTCGAAACA 91825
      :||| :||| :||| :||| :||| :||| :||| :|||
570 euAlaValArgValGlnTyTyrMetThrGlyLysPProSerGlnAlaLeu 586
      :||| :||| :||| :||| :||| :||| :||| :|||
91826 TTGCCGTTGGCTTCACATTAATGGCTGAGAAACCCCAAAAGATCGGCTC 91875
      :||| :||| :||| :||| :||| :||| :||| :|||
587 ProValIlePheGlyLysSerSerCysSerGluPheThrArgGluAlaPh 603
      :||| :||| :||| :||| :||| :||| :||| :|||
91876 AATGCCATCTACGCGAAGTCGCTTGTCCCAAAATTTTCTACGAAGCAT 91925
      :||| :||| :||| :||| :||| :||| :||| :|||
603 eThrProValValTyHisAsnLysSerProGluPheTyArgGluPheL 620
      :||| :||| :||| :||| :||| :||| :||| :|||
91926 TACAGCTGCAATTTATCATACAAAATGCCCATCATTTCTACGTAAGAAATA 91975
      :||| :||| :||| :||| :||| :||| :||| :|||
620 yLeuHisLeuProAlaCysValIThrGluAsnHisLeuLeuPheThr 636
      :||| :||| :||| :||| :||| :||| :||| :|||
91976 AATTTGCCGTGCGCTGCATTAAGACAGATCACCATTTATTTGTTACCC 92025
      :||| :||| :||| :||| :||| :||| :||| :|||
637 PheTyThrHisValSerCysGlnProArgPro.....GlyThrAlaLe 650
      :||| :||| :||| :||| :||| :||| :||| :|||
92026 ATTTCACAGTTTCAATGTCAAAAGAACACACAGACTTCAGCCGCTGAGT 92075
      :||| :||| :||| :||| :||| :||| :||| :|||
650 uGluThrProValGlyPheThrTyPleProLeuLeuGlnHisGlyArgL 667
      :||| :||| :||| :||| :||| :||| :||| :|||
92076 AAGAGCTCTTATCGGCTACACATGCTGCTGCTTACGAAAGTGAAGAAC 92125
      :||| :||| :||| :||| :||| :||| :||| :|||
667 euArgThrGlyProPheCysLeuProValSerValaspGlnProProPro 683
      :||| :||| :||| :||| :||| :||| :||| :|||
92126 TTAAGTTTGGGAATTTAATCTTCTGTATGTAAGTACACCGCGGAA 92175
      :||| :||| :||| :||| :||| :||| :||| :|||
684 SerTySerValLeuThrProaspValAlaLeuProGlyMetArgTyArgVa 700
      :||| :||| :||| :||| :||| :||| :||| :|||
92176 AATTTCTCATTTATACACAGCAATGTTCAACGCTCGGAATTAATATGCT 92225
      :||| :||| :||| :||| :||| :||| :||| :|||
700 lAspGlnHisLysGlyValAlaPheSerValGluLeuThrAlaValSerSer 717
      :||| :||| :||| :||| :||| :||| :||| :|||
92226 GGCACATTCACAGAGCGCGTGTCTTATTAATGTACAGACAGTACCGCAA 92275
      :||| :||| :||| :||| :||| :||| :||| :|||
```

```

717  aLhSPGlnAsPProTyrLeuAspLysPhePheThrLeuValHisVal 733
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92276  TYCAACTTTTGGAATCTCTTCCCTGCATGCATCTCTTAATTGGCAATAT 92325
774  LeuGIuGIuGIuAlaPhePProPheArgLeuLysAspThrValLeuSerGI 750
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92326  CTAGCACTCGAAATATATACCT.....TCCATATTGGCGA 92360
750  uGIyAsuValGIuGIuGlnLeuArgAlaSerLeuAlaLeuArgLeuA 767
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92361  GAATTAATATAGAAACAGAAATTAAAGAAATGCTGCTGCATATCGAATATG 92410
767  lAsPProGIuProLeuValAlaPheSerHisHisValLeuAspLysLeu 783
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92411  CAATACGTGACCCCTTATAGTGAGCATCTACCAATTATGTTGGATAGGTT 92460
784  ValArgLeuValAlaArgProPheIleSerGIuGlnIleValAsnLe 800
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92461  ATCGAATTTGATAGTTGTCACCCACAAAGTTGTTGGACAAAGAATGTCCT 92510
800  uGIyArgGIuAlaPheGIuAlaMetAlaHisValAlaSerLeuValHisA 817
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92511  GGGACCTACGGTTTTCGAAAGTCTTTGTTGGTACCACTTGTGTATACAA 92560
817  rGSerLeuGIuAlaAlaGlnAspAlaArgHisLysCysProGIuLeuAla 833
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92561  TACTTACAGCAGCATCATATGACCAATACGGTGGCAAAAGTTGCTATCA 92610
834  AlArYyValHisTYrAlaPheArgLeuProGIuThrGIuProSerLeuPr 850
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92611  ACATATGTGCACCTTTCATATGTAAATTCGCGA..... 92642
850  oAspGIuAlaProPheValThrValGlnAlaAlaPheThrLeuAlaArg... 865
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92643  .....CCCTTTCAAGCAACACAGCGACTATCAATGCGAGTGGC 92679
865  ..... 865
92680  AGCACAACAGAAAGACTGCATTAAGCAATGTCACACATATATGATTA 92729
866  .....GlySerGIuArgProAlaSerLeuTY 874
92730  TGTTTAGCAACAGTGGCGAAGCTTGGATCGCAAAAG...TAAATATACATA 92776
874  rLeuAlaArgSerLysSerLLeSerSerSerAsnProAspLeuAlaValA 891
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92777  TTTTGTAAAGGCTTAAATATATATACCTTTTCTTTCACAGAGCTGTCCATA 92825
891  lArProGIuSerValAspAspGIuValSerArgIleLeuAlaSerLysLeu 907
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92826  .....GAGTCTCTTCACTCGATGAGTGCACGGAGGATGCCAAGTGGCGGTG 92870
908  LeuHisGIuGIuLeuAlaLeuGIuThrArgValAlaSerSerSerAlaValAr 924
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92871  CTTCACGAGAGATTAAGCCTTGATTTGGGTTGTTGGCAAGTGGAAAGGCC 92920
924  gGIuAlaIleLeuGlnHisAlaTrpPhePheGlnLeuMetValLysS 941
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92921  TCATAGTAGCAATTCGCAAAATTCGTGGTTTTTATTTGAGCTTATGTCAAAT 92970
941  eTMeTAlaLeuHisLeuLeuLeuGIuGlnArgLeuAspThrProArgLys 957
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
92971  CTATGATTTGAAACATTTGCAATTTGTTCAAAATCTTTAAATGGTCCCGCAAA 93020
958  LeuArGpHeProGIyArgPheLeuAspAspIleThrAlaLeuValAlcYse 974
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
93021  CATCGATTTCCACATACAGTTTAAATATATATATCATATCCACCTGTGCTACTT 93070
974  rValGIuLeuGIuValIleThrArgValHisLysAspValGIuLeuValag 991
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
93071  ACTTACACAAATATTTGGTGGCTTACACAGCAACACACTTAAATTTGGAGG 93120

```

991	lUhlHSuSnaSLaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeu	1007
99121	AGCTCAATTAATCCAGATTAGAGCTTTTATATTGACATACATAGCATC	93170
1008	ValASpAGSLyPheValPheSerLeuValaGlaHisTyrLySGluVa	1024
93171	ATGGATCGGGGATTTGTTGTTGGCCCTCATCAAAAGCTACAAAGSTTCT	93220
1024	lAlaThrArgLeuGlnSerSerProAspProAlaLeuLeuThrLeu	1041
93221	GATTTCTAAAAATGCTCAATACCGGAT.....CTAATGAACATA	93261
1041	rGMeGLuPheThrArgTlIeLeuCysSerHisGluHisTyrValThrLeu	1057
93262	AAATAGACTTTTGTGGAATGTGTGTAGTACAGACATTTGTACCTTA	93311
1058	AsnLeuPro.....CysCysProLeuSerProAlaSerPr	1070
93312	AACTTGCCCTTTGGTACTTCATACACATGTATACAGCGCCATGCAGCC	93361
1070	oSerProSerValSerSerThrThrSerGlnSerThrPheSerSerg	1087
93362	CAGCGCAACACACAGCTGCAGACAGCCAAACTTCTCGTAAAGTAA	93411
1087	lAla.....	1088
93412	AAACATACACTTATTTTACTATTTGTGGCTTCTTAAAGGAT	93461
1089ProAspProLysAlaThrSerMetPheGluLeuSerGlyProph	1103
93462	CAGTACAGAGAGCCCTTCAGCC.....GACCTTAGTCAAGATT	93501
1103	eArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeu	1120
93502	CCGCCAGCAACATTTCTAGTGGACTTTCTGAGCGACTTAGACAGAG	93551
1120	lAlaGlu...ProGluAlaGluGlyAlaPheLeuLeuHisTyrSala	1135
93552	TAAATGAGGTGCCAAACCCCA.....TTACATGGAAGAAAGCA	93589
1136	lIleSerAlaValHisSerLeuLeuGlyHisAspThrAspProArgly	1152
93590	ATTAAGGTGTAATCGCAACGTGATGACATCTCAGATTGTGGAGCGGGTA	93639
1152	rAlaGluAlaThrValValLysAlaArgAlaGluLeuTyrLeuProLeu	1169
93640	TACCGAACTGATGCACGTGCAGAGGTGGCTGTTATATATACCGCTAC	93688
1169	eSerlIleAlaArgAspThrLeuProArgLeuHisAspPheAlaGluGly	1185
93690	TTTCTATAGTGAATGGATTGCATTCTCAGTTGCATCAGACAGGCTTGGAG	93733
1186	ProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThrGlu	1202
93740	CAGAGTCGTCTGCAGAAATCGGTCACTTGAAAGCATACAAAGGCCACA	93789
1202	yGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAlaIle	1219
93790	TCAAACTATTACAGCATCAACCTTAATCCGAGAGTGGCATTTGGCATAT	93833
1219	lAlaGlyProLeuAlaProGlySerArgAlaSerlIleSerGlnGlyPro	1235
93840	CGGGAAGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	93889
1236	ProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerSerArgTh	1252
93890	CCC.....CTCAGCTCGAAGTAATACCGGCA	93915
1252	rIleuLeuAlaCysValLeuThrValLeuLeuLysAsnThrGluProAlaLeu	1269
93916	TTTATTTGGTGTCTTCTGTGGGTCTTAAAAACCTTGAGCGCAACGTTT	93966
1269	eGlnAlaArgThrAlaThrAspLeuThrLeuProGlnLeuAlaGluLeu	1285


```

93966  ||||||| |||||: : : : : : : : : : ||
      TRPATCGTGGCTTATGGATTATTAAGTCCGGATGACGACCAATATCTT 94015

1286  AspleuLeuTyrLeuCysLeuAlaAlaPheGluTyrIleGlySerLys.. 1301
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94016  CAAGGTGTAACGTTTGCTGAACACTTTCAGTACACAGACAAAAA 94065

1302  .....AlaPheGluArgIleAsnSerLeuThrPheLysLys.....S 1314
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94066  CGTGGCCACCTTAACGAAACGATACCAAGTTTAGGAAACTGGAT 94115

1314  erLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrIleGly 1330
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94116  CTACTGATGTTAAAGAAAGCTGGAGAGATGATATAGAGGCACAAATTC 94165

1331  AlaArgGlnGluMetValArgArgSerArgGluArgSerProPheGlyAs 1347
      ||||| : : : : : : : : : : : : : : : : : : : : : :
94166  GCAGGTATGATTTAATAAATCGTCGAAAGATGTAAAT.....TC 94206

1347  nProGluAsnValArgTTPArgLys...SerValThrHisTrrPlyGlnT 1363
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94207  AACAGAAAAATTCGATGAGAGAAAGACAGATGCCATATCGCTCCAT 94256

1363  hrSerAspArgValAspLysThrLysAspGluMetGluHisGluAlaLeu 1379
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94257  ATGGGATGCTGCGGAAAAAGTACGATGACTGTAATTAAGTCATTTTC 94306

1380  ValGluGlnLeuLeuAlaThrGluAlaSerLeuValValLeuAspThrLe 1396
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94307  ATCGAAGGCTTTGGCACTGAAGTTCGACTTCTCTGATACCT 94356

1396  uGluIleIleValGlnThrValMetLeuSerGluAlaArgGluSerValL 1413
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94357  TGAATATTGTATGCTCTCT.....GCCAACCTTTATCAACACTTTC 94400

1413  euGluAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSer 1429
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94401  TTGAACGTGCTTAAGTCTTCTTCTTCTTCTTCTGCGGATTCATCT 94450

1430  AlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLy 1446
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94451  GTTTGGCCCTGCMAACTGTTGGCTCCCAACGCGCTTAATCTTTAA 94500

1446  sPheProGluLeuLeuPheGluGluAspThrGluLeuCysAlaAspLeuc 1463
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94501  ATTTCCAACTTATTATTATGATGACGACGCAATATATGCGGATTTGG 94550

1463  ySLeuArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThr 1479
      ||||| : : : : : : : : : : : : : : : : : : : : : :
94551  GTCTAATTCTACTGAACAATTGGGGCCCTATTGCCAGGAATACATCA 94600

1480  HisAlaSerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGly 1496
      : : : : : : : : : : : : : : : : : : : : : : : : : :
94601  CAGGCACGCTCATCATATATTACTAATGACACAAATTTTGAATTTGG 94650

1496  y..His..... 1497
      |
94651  AAATGTAGTATTATTATTATTAGTTTATTATTAAATTAATGCT 94700

1498  .....AsnPheAlaArgValLysMetGlnValThrMetSe 1509
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94701  CACTTATTCGTATAGAACTTGGCCGAGTTAAGATGCAAGTGCAGTTC 94750

1509  rLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGluHisLeuA 1526
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94751  TTTTACGTCCTAGTTGAACAAGTTCGGCTTTAGTGACAAATCCTTTC 94800

1526  rArgSerSerLeuTyrThrIleLeuThrTyrAlaGluGluAspMetGlyLeu 1542
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94801  GCCGTGGCTTAAACGCTGCTGGTTATGTAATCCGATCCGACGCTGG 94850

1543  ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHi 1559
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

94851  CAGGATACTCTTTTCCGAAACAGTACAAGATTTCTTTTAATTCGCA 94900

1559  sMetIleLeuThrAspThrValLysMetLysGluHisGlnGluAspProG 1576
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94901  TATGATCTCTGTGAGATCTGTTAAATGAAGAGATTCAGGAAGCCAG 94950

1576  LuMetLeuIleAspLeuMetTyrArgIleAlaArgGlyTyrGlnGlySer 1592
      |||||||: : : : : : : : : : : : : : : : : : : : : :
94951  AAATGTGCTTGACCTCATGAATCGATTTGCCAAGGATACCAAAATAC 95000

1593  ProAspLeuArgLeuThrTrrPheGlnAsnMetAlaGlyLysHisAlaGly 1609
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95001  CCGATCTTACGACTGACTGTTGGAAAAATATGGCTAAAAAAACACCGCA 95050

1609  uLeuGluAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaL 1626
      | : : : : : : : : : : : : : : : : : : : : : : : : : :
95051  GCGAGCAATTCACACGGAAGCAGCCATGTTATGATACATGCTGCTTCT 95100

1626  euValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHisLeuProVal 1642
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95101  TAGTTCTGAATATCTTACGATGTTGAGTCCAAAAACATTTGCCGCTT 95150

1643  GlyCysValSerPheGluAsnIleSerSerAsnValLeuGluGluSerAl 1659
      |||: : : : : : : : : : : : : : : : : : : : : :
95151  GAGGCTGATAGTTTTCACGAATTTCTCCACACACATTAATGAGTCCGC 95200

1659  aIleSerAspAspIleLeuSerProAspGluGluGlyPheCysSerGlyL 1676
      | : : : : : : : : : : : : : : : : : : : : : : : : : :
95201  CGTATCGGATGATGTGCTAAGTCCCGGCAAGATGTTCTGCCATAGGAA 95250

1676  yHisPheThrGluLeuGlyLeuValGlyLeuLeuGluGlnAlaAlaGly 1692
      : : : : : : : : : : : : : : : : : : : : : : : : : :
95251  ATCATTTTCACTAAACTGGGTGAAGCCCTTGCGAAGAACCTCCAAAT 95300

1693  TyrPheThrMetGlyGlyLeuTyrGluValAlaGlnGluValTyrLysAs 1709
      || : : : : : : : : : : : : : : : : : : : : : :
95301  TCTTTTCAAGTTGCTGCGCATGATATGACGACATGAACGAAGTGTCAAAAT 95350

1709  nLeuIleProIleLeuGluAlaHisArgAspTyrLysLysLeuAlaVal 1726
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95351  TCTAATACCATATATCGAGGCTTAACGAGATTTCAAAAGCTAAGCAAG 95400

1726  aHisGlyLysLeuGlnGluAlaPheThr..... 1735
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95401  TTCATGCAAAATTCGAGAGGCAATTAATCGAATATCCCACTACAGGTA 95450

1736  .....LysIle..MetHisGlnSerSerG 1743
      ||||| : : : : : : : : : : : : : : : : : : : : : :
95451  ACAATATTGTGAATTTTACCACGCAAAATATATATCATATTTATTAAC 95500

1743  lYTrpGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAla 1759
      : : : : : : : : : : : : : : : : : : : : : : : : : :
95501  AGGTAAGAGATTTTTGAACATATCTTCTGTTGGCTTATAGGGGGA 95550

1760  HisPheGlyAspLeuAspGluGlnGluPheValTyrLysGluProSerIl 1776
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95551  AAATTTGGGACTTGGATCAGCAGGAATTCATTATTAAGGCCAACATY 95600

1776  eThrLysLeuAlaGluIleSerHisArgLeuGlu..... 1787
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95601  CACGAAGTTCGCCGAATATTTAGTGGCTTCAAGTATATATTGCAAAAT 95650

1788  .....Glu..PheTy 1790
      |||||||
95651  GGAATAATATAGACAAATCAATTTTGTTCACAACTATAGTTAAGACTTTA 95700

1790  rThGluArgPheGlyAspAspValValGluIleIleLysAspSerTyrP 1807
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95701  CACTGAAGATTCGACCGGACTGTGATATCATTAAGATTCACATA 95750

1807  roValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThr 1823
      |||||||: : : : : : : : : : : : : : : : : : : : : :
95751  CCGTTGATATTAAATAGCTTGATCCGATTAAGGCTTACATTCAATTAAT 95800

```

```
1824 TyrValGluProThrPheAspThrTyrGluLeuLysAspArgValThrTy 1840
|||||
95801 TATGTTGAACCTTACTTGTGAACATATGAAATGCGTATCGTAGACATA 95850
|||||
1840 rPheAspArgAsnTyrGlyLeu..... 1847
|||||
95851 CTTTGACGGCAATTTTCATATATAGTATGATATGAATTAACAGATAATT 95900
|||||
1848 ..... 1849
|||||
95901 AAATCGAATTTTAAATATATTTAACTTTTGGTTTATGCAGACAGC 95950
|||||
1850 PheLeuPheCys.ThrProPheThrProAspGlyArgAlaHisGlyGlu 1866
|||||
95951 TTTTATATATAGCCACGCGCTTTACTTAAATGGAAGGCGCATGCGCAAC 96000
|||||
1866 euProGluGlnHisLysArgGlyThrLeuLeuSerThrAspHisAlaPhe 1882
|||||
96001 TAAATGACGCGTGAACGGAACGATTTTGCACCGCGGCAATCACTTT 96050
|||||
1883 ProTyrIleLeuThrArgIleArgValCysHisArgGluGluThrValIle 1899
|||||
96051 CCATAGCGTAAACCTCGCATTTATGTAATGAGCCAGACGCAATAGTCTCT 96100
|||||
1899 uThrProValGluValAlaIleGlu..... 1907
|||||
96101 GGAACCGATGAGTGGCATGTGAAGTATGTTGTAAATCCCAAA 96150
|||||
1908 ..... AspMetGlnLysThrArg 1914
|||||
96151 AGTTAATAATTAACGTTATTGCTAGATATTCATAAAAAACAATTG 96200
|||||
1915 GluLeuAlaPheAlaThrGluGlnAspProProAspAlaLysMetLeuG 1931
|||||
96201 GAATGCGACGTCGACAGATCAAGAGCCAGACGCCAATAATTTGCA 96250
|||||
1931 nMetValLeuGlnGlySerValGlyProThrValAsnGlnGlyProLeuG 1948
|||||
96251 AATGGACTCAGAGGATGATGTGAAACCACTGTTAACCAAGACGATGG 96300
|||||
1948 lValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeu... 1963
|||||
96301 AATGGGAGATGTCTCTTCCATTTATCCAGCGAACACTGTACCG 96350
|||||
1964 PheArgHisHisAsnLysLeuArgLysPheLysAspPheCysLysTy 1980
|||||
96351 ACAAAACACCAAAACAATCTTCGTTGCTTCGCGAGTTTCAAAAGCG 96400
|||||
1980 scyGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGln 1997
|||||
96401 TTGTGCTGAGGCTTTGAAAAGATCGCAATCTTAAGACTTTCAGATCAA 96450
|||||
1997 ysgLysTyrHisArgGluLeuGluArgAspTyrLysArgLeuGluAla 2013
|||||
96451 AAGATTACCAACAGACTGGAACGTACACAGATGGTTCAATTGAACGA 96500
|||||
2014 LeuGlnProLeuLeuThr 2019
|||||
96501 TTGACTCCCTTTATTACT 96518
|||||
seq_name: gb_pr:BC008335
seq_documentation_block:
LOCUS BC008335 2825 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, clone IMAGE:3506202, mRNA, partial cds.
ACCESSION BC008335
VERSION BC008335.1 GI:14249910
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
AUTHORS 1 (bases 1 to 2825)
          Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (25-MAY-2001) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
          Email: gcgaps@email.nih.gov
          Tissue Procurement: DCTD/DTF
          cDNA Library Preparation: Rubin Laboratory
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          Info@gsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
          Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Letlicia Hsiao, Martin Krzywnicki, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderduyn, Marco Marra.
FEATURES
source Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
          Series: IRAL plate: 23 Row: d Column: 12.
          Location/Qualifiers
            1..2825
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:3506202"
              /tissue_type="kidney, renal cell adenocarcinoma"
              /clone_id="NIH_MGC_14"
              /lab_host="DH10B-R"
              /note="Vector: pOTB7"
            <1..2585
              /codon_start=3
              /product="unknown (protein for IMAGE:3506202)"
              /protein_id="AAH08335.1"
              /db_xref="gi:14249911"
              /translation="ORSRLASMDSPTEBEGDIAGINPVSAMATAGSPAGSRASI
              SGGPPTASRAGCLSAESSRTILACTVYVTKTEPDLDRMAITDITLPUGLLDILY
              LCLAAFEYKKTKAFERINSLTEFKSLDMKARLEALIGTGAEQWVRSRSPGN
              PENVRRKSVTHKKQTSRDYKTKDMEHEALVEGLATVASLVDTEILVQVYM
              SEARESYGAVLKVVLVYLSGASQSAFLQGLATQALVSKPELLEFDETELDLC
              RLRLRHGSRISIRITHASASLYLAKRONPEIGHNARVAKMOVNMLSSLVGTQNF
              EEHRLRSIKITLTYAFEDMCIKDSPTFAEYQDLMFIMHILINDVYMKRHOBDPMLI
              DMWRLARGTQSPDLRLTYLQNMAKNHELDGHNHAFACQVHAAALVAVETALAEDEH
              RHLPEVCVSFQNTSSNVLESALISDILSPDEGFCGSHFTELGLVLEQAGFT
              MGILYAVNVEVYNNLIPILEAHNDYKKLAIVGKLOEARTKIMHOSGWERVGYFR
              VGFYGAHFGDLEDEFEYKESITTKLAETSHRLEEFYTERPDDVVEIKDNPYKS
              KLSOKAYOIOTVVEPYPDYELKDDYEDRNRYGLRFTLFCPTPPDQRAAGELPEO
              HKRKTILSTDHAPYIKTRIRVCHRETYLITPEVAIEEMOKKTEPLAATQDDPDA
              KMLQNVYQGSVGPPTVQNGPLEVAQVFLAIPEDPKLFRHNHLRUCFDQCKCEDAL
              RKNKALIGDPQKEYARELERENYCRLEALQPLTORLPQJMAETPPGLKNSLNRSFR
              KADL"
BASE COUNT 641 a 835 c 827 g 522 t
ORIGIN
alignment_scores:
Quality: 4419.00 Length: 860
Ratio: 5.144 Gaps: 0
Percent Similarity: 99.884 Percent Identity: 99.767
alignment_block:
US-09-736-968a-2 x BC008335 ..
Align seg 1/1 to: BC008335 from: 1 to: 2825
1188 GlnArgSerArgLeuAlaSerMetLeuAspSerAspThrGluGluGluGlu 1204
```

|||||
3 CAGCGGTCAAGACTGGCCCTCAATGCTTGACTGACACACAGAAAGCGAAGG 52
1204 YASPIIAIAGLYThrIleAsnProSerValAlaMetAlaIleAlaGLYC 1221
53 GGACATGCGGGGThCAATCAACCCCTGTGGCCATGGCCATTGGTG 102
1221 LProLeuAlaProGLYSerArgAlaSerIleSerGlnGLYProCotr 1237
103 GCCCCTAGGCCCGTGGCTCCGGGGCAGATCTCCAGGGGGCCACCAAG 152
1238 AlAserArgAlaGLYCysAlaLeuSerAlaGluSerSerArgThrLeu 1254
153 GCTTCTGCGCGAGGCTGTGCCCTCTCTGAGTCAAGCCGAGCTTGCT 202
1254 uAlaCysValIleuThrValIleuYAsnThrGluProAlaLeuGln 1271
203 GCGCTGTGCTGCTGGGTGCTGA AAAACACCGAGCCGGCGCTCTCAGC 252
1271 rGTTrAlaThrAspLeuThrLeuProGlnLeuGlnArgLeuAspLeu 1287
253 GGTGGCCACTGACTGACTGACTCCCGAGCTGGAGCTGTGTGGACTTG 302
1288 LeuTYrLeuYsLeuAlaAlaPheGluTYrLYSGLYLysLYSAlaPheG 1304
303 CTGTACCTTTGGCTGGCTGCTTTGAGTACAAAGGGAAAAAGCCTTGA 352
1304 uArgIleAsnSerLeuThrPheLYSLeuYsSerLeuAspMetLYSAlaArg 1321
353 ACCGATCAACAGCCTCACAATTCAAAAATCTCGATATGAAAGCCCGGC 402
1321 euGlnGluAlaIleLeuGLYThrIleGlyAlaArgGlnGluMetValArg 1337
403 TAGAGGAAGCCATTCTGGGTACCATCGAGCTGACAAAGAAATGTTCCG 452
1338 ArgSerArgGluArgSerProPheGluYAsnProGluAsnValArgTrpArg 1354
453 CGAAGGTGTAGAGAGAGCCGTTTGGAAATCCAGAGAACGTGGCTGGGG 502
1354 gLYSserValThrHisTrpLYSGlnThrSerAspArgValAspLYSThrL 1371
503 GAAGAGGCTACACACTCGAAGCAAACTCAGACCCGCTGGACAGACCA 552
1371 yAspGluMetGlnHisGluAlaLeuValGluGlyAsnLeuAlaThrGlu 1387
553 AGGATGAATGGAACACGAGGCTTGTGTGAAGGAACCTGGCAACCGAG 602
1388 AlAserLeuValIleuAspThrLeuGlnIleIleValGlnThrValMet 1404
603 GCAAGCCTAGTGGTCTGGACACACTGGAGATCATCTGCAGACGGTGAT 652
1404 tLeuSerGluAlaArgGluSerValIleuGlyAlaValIleuLYSValVal 1421
653 GCTTTCGAAGCCCGGAGAGGCTCTTGGGGGAGTGTGAAGGTTGTGC 702
1421 euTYrSerLeuGlySerAlaGlnSerAlaLeuPheGlnHisGlyLeu 1437
703 TGTACAGCCTGGGCGAGTGGCCAAAGTGCCTTCTTGCACAGATGCCCTG 752
1438 AlaThrGlnArgAlaLeuValSerLYSPheProGlnLeuLeuPheGln 1454
753 GCCAACCAAGGGCCCTTGTGTCCAAGTTCGCCGAGCTGCTGTCCAGGA 802
1454 uAspThrGlnLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysG 1471
803 GGACAGGAGGCTGTGTGCCAGCTGTGCCGTGAAGCTCTTACGACACTGTG 852
1471 LysArgGlnSerThrIleArgThrHisAlaSerAlaSerLeuTYrLeu 1487
853 GCAGCGGCATCAGACCATCCGACACGACGCGCTGCTGTAACCTG 902
1488 LeuMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLYSMe 1504
|||||

903 CTCATGCGACAAACTTCAGATCGGCCACAACATTGGCCCGTGAAGAT 952
1504 tGlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPhe 1521
953 GCAGGTCAACATGTCTCTGCTCCCTGTGGGGAGCAGCAGAACTTCA 1002
1521 euGlnGlnHisLeuArgArgSerLeuLYSThrIleLeuThrTYrAlaGlu 1537
1003 GTGAAGACACACTGGGAGGTTCACTCAAAACCATCTCACTATGTGAG 1052
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLe 1554
1053 GAGGACATGGGGCTGCGGAGACGACCTTCCAGAGCGAGTCCAGGACT 1102
1554 uMetPheAsnLeuHisMetIleLeuThrAspThrValLYSMetLYSGlu 1571
1103 GATGTTCAACCTGCACATGATCTTGACGAGACAGGTGAAGATGAAGAAC 1152
1571 LsGlnGluAspProGluMetLeuIleAspLeuMetTYrArgIleAlaArg 1587
1153 ACCAGAGGAGCCCTGAGATGCTCATCGACSTCATGTACAGAAATGGCCGG 1202
1588 GLYTYrGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAl 1604
1203 GGTTCACAGGGCTCACCGACCTTCGCTGACTGTGTGCAAACTATGGC 1252
1604 aGlyLYSHisAlaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetV 1621
1253 CGGAAAGCACCGGAGCTGGGGCAACCCAGGAGCCGCCAAGTGCATGG 1302
1621 aHisAlaAlaAlaLeuValAlaGluTYrLeuAlaLeuLeuGluuAspGln 1637
1303 TGCACGCGCGCCCTCGCTGGCTGAGTACTCGCCCTGCTGAGAGACAC 1352
1638 ArgHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnVal 1654
1353 GCGCACTGCGCGGTGGCTGGCTTCTTCCAGAACATCTCATCAACGT 1402
1654 LLeuGlnGluSerAlaIleSerAspAspIleLeuSerProAspGlnGlu 1671
1403 GGTAAAGGAGTCCGCAATCTCGAAGACATCTGTCCGCCACAGAGAGG 1452
1671 LYPheCysSerGlyLYSHisPheThrGluLeuGlyLeuValGlyLeuLeu 1687
1453 GCTTCTGCTCCGGGAGCACTTCACTGAGCTGGGGGTGGTGGTCTCTG 1502
1688 GlnGlnAlaIaGlyTYrPheThrMetGlyGlyLeuTYrGluAlaValAs 1704
1503 GAACAGGACCGCGGTACTTCAACATGGGCGGCTCTTACGAGGGGTGA 1552
1704 nGluValTYrLYSAsnLeuIleProIleLeuGluAlaHisArgAspTYrL 1721
1553 TGAGGTCTCAAAAGAACCTCATCCCATCTCGAAGCCACCGTACTACAA 1602
1721 yLYSLeuAlaIaValHisGlyLYSLeuGlnGlnAlaPheThrLYSile 1737
1603 AGAAGCTGGCGCGGTGCACGCAAACTGCGAGGAGGCTTACACAAATC 1652
1738 MetHisGlnSerSerGlyTYrPGLuArgValPheGlyThrTYrPheArgVal 1754
1653 ATGCACCAAGATTCGGGTGGAGCGGCTGTTCGGAGCGATTTTCCCGGT 1702
1754 LGlyPheTYrGlyAlaHisPheGlyAspLeuAspGlnGlnGluPheValT 1771
1703 GGGCTTACAGCGGCCCACTTGGGTGACCTGGATGACAGAGAGTTTGTGT 1752
1771 TYrLYSGluProSerIleThrLYSLeuAlaGluIleSerHisArgLeuGlu 1787
1753 ACAAGAGACCATCATCAGACGAACTGGCAGAGATCTCACACCGGCTGAG 1802
1788 GluPheTYrThrGluArgPheGlyAspAspValValGluIleIleLYSAs 1804
1803 GAGTTCTACACGAGAGATTGGCGAGAGAGTGTGATGATTATCAAGA 1852
|||||

seq_name	gb_pat	AX172980
seq_documentation_block	4200 bp	DNA
LOCUS	AX172980	linear
DEFINITION	Sequence 107 from Patent WO0142297.	PAT 03-JUL-2001
ACCESSION	AX172980	
VERSION	AX172980.1	
KEYWORDS		
SOURCE	synthetic construct.	
ORGANISM	artificial sequence.	
REFERENCE	1 (bases 1 to 4200)	

[illegible]

597	CTCAATATCAAGCCCTGTCATTTCTTTCCAAATGATCTGGTGTGTTAATGGA	646
1009	parGAlpPhvetaIpheserLeuValArgAlaHisTyrLysGlnValAlat	1022
647	CAGAGGATTTGTTTAAAGTCCCTGTATTAACGAGGTGCTT	696
1026	hrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMet	1042
697	CAAAAGCTTTACTCATTTACCAAAATCCCAAGGTTCTGGTGTCTTGAAGGCTG	746
1043	GlnpHrhrArgLeuLeuLeuLysSerHisTslGlnHisTyrValAlThrLeuAsnLe	1053
747	GAATTTTGTACGAATCAATCTGCATCTCTGAGCACTATGTTAACTTAACCTT	796
1059	uProCysCysProLeuSerProProAlaSerProSerProSerValIserS	1070
797	ACCTGAGGCTTACTTACCTCAACCTGCATCTGCATCACCCTTGTGTTCTT	846
1076	erThrThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLys	1092
847	CTGCACAACTCTCAGAGCTTCGATTTTTCACGAATGTACAAAGACCAAAAG	896
1093	ValThrSerMetHrheGluLeuSerGlyProPheIleArgGlnGlnHisPheLe	1109
897	ATTGCAAAATATGTTTAAATTTATCCGGCTTTTCCGCCAAACAGCATTAATTT	946
1109	uAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlnuProGluuAlaG	1126
947	GGCAGGACTTGTGTTAACAGAGCTGGAGCTGTCTTTTAAACCTGTAACTGTG	996
1126	uGluGlyAlaPheLeuLeuHisTslLysAlaAlaSerAlaValHisSerLeu	1142
997	AAGGACTGTTTGATTTGCATTAAGAAAGCATCAATATGTACACAACTTTA	1046
1143	LeuCysGlyHisAspHrhrAspProArgTyrAlaGluAlaThrValLysAl	1159
1047	CTCTCCAGCTACAGACACACACCCTGGATCTGTGACCTCCAGATTAAGAGGC	1096
1159	aaArgValAlaGluLeuTyrLeuProLeuLeuSerIleLeuIleArgAspHrL	1176
1097	TCCAGTGGCCATGTGTATCTACCTCTGTATTTGGTATTTTCATGGAACGTG	1146
1176	eudProArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeu	1192
1147	TACCTCAGCTGATGTGATTTTACAGAAACATCAACATCAACGAGGAAGACA	1196
1193	AlaSerMetLeuAspSerAspHrhrLysGluGlyAspIleAlaGlyTh	1209
1197	ATTTGTATGACCACTGATATATTATTAAGGTGAGAC.....GGAGTAT	1240
1209	rIleAsnProSerValAlaMetAlaIleAlaGlyGlyProLeu.....	1223
1241	GATTAACCAAGACCGTTGGCATGTGGCAATGGCAGGAGCATCGGATCCCTCAAC	1290
1224AlaProGlySerArgAlaSerIleSerGlnGlyProProThrAla	1238
1291	TAAACAAGCCCTGGACAGTTCTCTCTCACAGTCA.....ACGAGT	1328
1239	serArgAlaGlyCysAlaLeuSerAlaLeuSerSerArgHrhrLeuLeuAl	1255
1329	GGAGGCAACACACTAACCTTTTTCACACAAATTCAGTCCAAACCCCTTTGAT	1378
1255	acArgValLeuThrPValLeuLysAsnThrGluProAlaLeuLeuGlnArgT	1272
1379	CTGTCTACTTTGGGTTTCCAAAATATGCAGATGAACAGTTTACAGAAAGT	1428
1272	rPalaThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuLeu	1288
1429	GGTTTACAGATCTCTCAGACGCTTGACAGTAACACGGATATTAGATCTGCTT	1478
1289	TyrLeuCysLeuAlaAlaPheGluTyrTyrLysGlyLysLysAlaPheGluAr	1305
1479	TATCTCTGTGTGTCTGTGTGAGATTAAGAGGAAAAAAGTTTGAACG	1528

1305	gllhnsrSerLeuThrPheLysLysSerLeuaspMetLysAlaArgLeu	1322
1529	AATGAATAGCTTACCTTTAAGAAATCAAAAGACAGAGCAAGGACTTG	1578
1332	lucLualAlaLeuLeuGlyThrIleGlyValArgGlnGluMetValArgTrg	1338
1579	AAGAAGCTATTCTTGCGAGCATAGGCGCCAGGCAGAAATGGTACGGCGA	1628
1339	SerArg.....GluArgSerPro.....PheGlyAsnPr	1348
1629	AGCCGAGACAGCTCGAGAGAAACCCATCTGGAAAGCTTTGGAAAGTCA	1678
1348	ogLAsnValArgTrpArgLysSerAlaIlnhrIsthrPyrLysGlnTrpSerA	1365
1679	AGAAATTTTGAAGTGGAGGAAAGATATGACTACAGGGGTCAAAACACAG	1728
1365	spArgValAspLysThrLysAspGluMetGlnHISGlnAlaLeuValGln	1381
1729	AGAAAGCTTGCAATATCAAGACGACAGATTTGAACACGAAAGCACTGATT	1778
1382	GlyAsnLeuAlaThrGlnAlaSerLeuValValLeuAspThrLeuGlnTrl	1390
1779	GGAAACCTGGCTACAGAACCAACCAATCATTTATGATTAATAGAGAT	1828
1398	eileValGlnThrValMetLeuSerGlnAlaArgLysSerValLeuGlnGlyA	1415
1829	TGTTGTTACAGACCTGTTTCTGTACAGGAATCCAAAGAGCAATCTTGGTG	1878
1415	lavalLeuLysValValLeuTrpSerLeuGlySerAlaGlnSerAlaLeu	1431
1879	GAGTGCTAAAGAGCTACTACACAGCATGGCTGTACCAAAAGTGCACTT	1928
1432	PheLeuGlnHISGlyLeuAlaThrGlnArgAlaLeuValSerLysPhePr	1448
1929	TATCTACACACACTGTTTCTCTACACAGAGGCTTGTTTAAAGTTTCC	1978
1448	ogLLeuLeuPheGlnGluAspThrGlnLeuGlyAlaAspLeuGlySerLeuA	1465
1979	TGAACCTTATTATGAAGAGACAGACAGAGAGTGCTGATTATGCGTCA	2028
1465	rgfLeuLeuArgHISGlySerAlaGlyIleSerThrIleArgThrHISAla	1481
2029	GGCTTCTCCGACACTGTACACATTCGATACCGATCAATCGGTGCACACCC	2078
1482	SerAlaSerLeuTrpLeuLeuMetArgGlnAsnPheGlnIleGlyLysAs	1498
2079	AGAGCTCCCTTACTACTACTATATGAGGCAAACTTGAATGGGAATTA	2128
1498	nPheAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuValG	1515
2129	CTTTGGCAGAGGTTAAATATCAGAGTACCAATGCACTATCCCTCTGGTGG	2178
1515	LyrThrGlnAsnPheSerGlnGlnHISLeuArgArgSerLeuLysThr	1531
2179	GCACATCTAGAAATTTTATATGAAGAATTTCTTAAGCGTCTCTTAAAGCT	2228
1532	IleLeuThrTrpAlaGlnGlnAspMetGlyLeuArgAspSerThrPheAl	1548
2229	ATATTGACATATATGCGAAGAAATCTGGAATTTGAGGGCAACACATTTCC	2278
1548	agLuglnValGlnAspLeuMetPheAsnLeuHISMetIleLeuThrAspTr	1565
2279	TGATTCAGGCTCCAGGATCTGGTTTTCAATCTCCATATGATTTCTTCATTA	2328
1565	hryAlaLysMetLysGlnHISGlnLysAspProGluMetLeuIleAspLeu	1581
2329	CTGTGAATATGAAGAACACACAGAGAGATCTCTGAATATGTTGATGACTTA	2378
1582	MetTrpArgIleAlaArgGlyTrpGlnGlySerProAspLeuArgLeuThr	1598
3379	AGTATAGAAATTTGCCAAGGTTATCCAGACACCTCTCAAGATCTGGCATTTGAC	2428

1443 TTGAGTATTAAGGCAAAAAGTGTGTAACGATGATGACCTT 1492
1312 LysLysSerLeuAspMetLysAlaArgLeuGluGluAlaLeuGlyTh 1338
1493 AAGAAATCAAAAGCATGAGAGCAAAAGCTTGAAGACTATCTTGGAG 1542
1328 rTLeGlyAlaArgGlnGluMetValArgArgSerArg.....Glu 1342
1543 CATTAGTCCAGGCAAGAAATGGTACGGGAGGCGAGACAGCTCGAGA 1592
1342 rGSerPro.....PheGlyAsnProGluAsnValArgTrpArg 1354
1593 GAAGCCCATCTGGAGAGTGCCTTGGAAATCAAGAAATTTAGGTGAGG 1642
1355 LysSerValThrHisTrpLysGlnThrSerAspArgValAspLysTrpLys 1371
1643 AAAGTAATGACTCACTGGCTCAAAACACAGAGAAGCTTGCAAAATCAAG 1692
1371 sAspGluMetGluLeuGluAlaLeuValGluGlyAsnLeuAlaThrGlu 1388
1693 AGCAGAGATTGACACGAAAGCATGATTGATGAAACCTGGCTACAGAG 1742
1388 LaSerLeuValValLeuAspThrLeuGluLeuLeuValGlnThrValMet 1404
1743 CAACCTTAATCATTTAGATACATTAGAGATTGTGTTCAACACCTTCT 1792
1405 LeuSerGluAlaArgGluSerValLeuGluValAlaLeuLysValVal 1421
1793 GTAAAGGATTCGCAAGAGACGATCTTGTTGAGAGCTTAAAGAGCTACT 1842
1421 uTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeu 1438
1843 ACACAGCATGGCTGTAAACCAAGTGCAGATTACTACACACAGCTTTTG 1892
1438 LaThrGlnArgAlaLeuValSerLysPheProGluLeuLeuPheGlu 1454
1893 CTACACAGAGAGCCTTGTTCAAAAGTTCTCTGAACTTATTGAAAGAA 1942
1455 AspThrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysG 1471
1943 GAGCAGAGCGAGTGTGCTGATTATGCTCAGGCTCTCCGACACTGTAG 1992
1471 ySerArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyrLeu 1488
1993 CAGTAGCATGGTACATACGATGACACCCAGAGCCTCCCTTACCTAC 2042
1488 euMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504
2043 TAATGAGGCAAAACTTTGAGATTGGAAATACTTGGCAGGTTAAATG 2092
1505 GlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPhe 1521
2093 CAGGTACCAATGTCCATCATCTCTCTGTTGGGACATCTCAGAAATTTAA 2142
1521 rGluGlnHisLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGlu 1538
2143 TGAAGAAATTTAAAGCTTCTTAAAGACTAATGACATATGCTGAAG 2192
1538 LysAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLeu 1554
2193 AAGATCTGGAATTGAGGAAACAACATTTCTCTGATCAGGTCCAGATCTG 2242
1555 MetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGlnHis 1571
2243 GTTTTCAATCTCATATGATTCTTTGTGATACGTGAATGAAGAGAC 2292
1571 sGlnGluAspProGluMetLeuLeuLeuAspLeuMetTyrArgIleAlaArg 1588
2293 CCAAGAGAGATCTGAATGTGATTGATCTAAATGTACAGAAATGCCAAG 2342
1588 LysTyrGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAla 1604
2343 GTTACCAAGACTCTCCAGATCTCGATTGACTGGTTGCGAGAACATGGCA 2392

1605 GlyLysHisAlaGluLeuGluGlnHisAlaGluAlaAlaGlnCysMetVal 1621
2293 GGCAGACACTCAGACAGCAACATCATGCTGAAGCTGCACAGTGTACT 2442
1621 HisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuLeuGluAspGln 1638
2443 CCAGCTCAGCAGCACTTGTCTGAATATTTGAGCATGTGAGAGCCGGA 2492
1638 rHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnVal 1654
2493 AATATCTCTGTGGAGTGTACATTTCAAGAAATTTTCATCTAATGTT 2542
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGlu 1671
2543 TTGGAAGAATCTCGGCTCAGATGATGTGTCTCCAGATCAAGAAAG 2592
1671 yPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeu 1688
2593 TATCTGCTGTGAAATACTTTACTGATCAGGACTTGTGGATTTACTGG 2642
1688 LysGlnAlaAlaGlyTyrPheThrMetGlyLeuTyrGluAlaValAsn 1704
2643 AACAGACAGCTGCTTCTCTCTATGCTGCGATGTATGAGCAGTTAAT 2692
1705 GluValTyrLysAsnLeuLeuProIleLeuGluAlaHisArgAspTyrLys 1721
2693 GAAGTTTACAAATCTATTTCTCATTTACAGAACTAATCGGATGCAAA 2742
1721 sLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIle 1738
2743 GAACATATCCACCAATTCATGTAACTTCAAGAGCATTCACCAAAATG 2792
1738 eHisGlnSerSerGlyTyrProGluArgValPheGlyThrTyrPheArgVal 1754
2793 TTATATAGATGACTGCTGCGAGCGGATGTTGGCACTATTTTGTGCTT 2842
1755 GlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGluPheValTyr 1771
2843 GGTTTTATGAAACCAAGTTCGGGATTTGGATGAAACCAAGATTTGTTA 2892
1771 rLysGluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGlu 1788
2893 CAAGGAGCCTGCATTAACCAACTTSCAGAGATTCACAGATTGGAGG 2942
1788 LysPheTyrThrGluArgPheGlyAspAspValValGluIleLysAsp 1804
2943 GATTTTACGAGAAAGATTTTGAAGAGATGTGTTGAAGTAATCAAAAGC 2992
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG 1821
2993 TCTAATCCGTGAGACAGTGAATTAATTAAGTCTTACAAAGCATATTTCA 3042
1821 nIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArg 1838
3043 GATTAACCTATGTGAGCCATCTTGAACACATATGATGAAGAGCAGAA 3092
1838 aLThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854
3093 TCACCTATTTGACAAAAATTAACAATCTTGTGATTCATGTACTGTACA 3142
1855 ProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisLys 1871
3143 CCTTTACTTTGATGCGCGGCCCATGGGAGACTTCAATGAACATTTCA 3192
1871 sArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThr 1888
3193 AAGGAAGACCATTTGACAGTGCATGCTTCTTCTATATTAATAAACA 3242
1888 rGluArgValCysHisArgGluGluTyrValLeuThrProValGluVal 1904
3243 GGTCAATGTACTCATTAAGAAAGATCATCTTAACCAACCAATGAAGTT 3292

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="MESAN1000101"
 /cell_type="normal mesangial cells (NHMC56046-2)"
 /clone_lib="MESAN1"
 /note="cloning vector: pME18FL3-primary culture, normal.
 mesangial cells"
 54..2279
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAB71042.1"
 /db_xref="GI:16550750"
 /translation="MEPLSVPRROOHYLAGVLTSLAVITLDPAGEFLGKHKVNW
 HILSHSDPRIDPQIKARVAMLYPLGITIMEYLPQIDTEHNOGRICAT
 DYESSGSMISQTVAMALVSTVPLQRLRPSFLITSGRQHTTESAESRLTCL
 LVMKNADETVLQKMTDLSTVLRLDLILYCSCFEYKGRKVEPKMSLPKRSK
 DMRALKEALIGISGARQEWRSRQGLERSPSGASQSEDLRMRKDMHMQNTEK
 LKRSALIEHGLIDGNLATFANLIIDTLEIVYQVTSYTESKESIIGVLYKLHSM
 ACNOAVYLQHCFAFQORALVSKPELLEETEECCALCLRLRHSSTIGTIRSHAS
 ASVLYMRQNEFIGNNFARVMQVMTSLISVGSQSENEPILRSIKTILTYAEDL
 ELRTPPDQVODLYPNIHMTLSPTVYKMEKEDPEMLIDLMTRIAKYOTSPDLRT
 WIQMAGKHSRSHNAEQCLHSAALVAEYLSMLDRKTYLPGCVTFQNISSNLE
 ESAVSDVYSPDEBICSGKFTESGLVGLLEQAAASFSAAGMEVAENEYKVLPIH
 EAMRDAKTLSTHGLQEAFLKIVHODGRMFQYFVGVGTGFGDLQDEHVEYEP
 AITKLAELSHRLSEFGEFGEFEDVEYIKDSNPYDKCLDPNTHIFLPMWSHTLHM
 R"

BASE COUNT 1046 a 640 c 711 g 930 t
 ORIGIN

alignment_scores:
 Quality: 3626.50 Length: 982
 Ratio: 4.130 Gaps: 7
 Percent Similarity: 89.409 Percent Identity: 71.996

alignment_block:
 US-09-736-968a-2 x AK055905 ..

Align seg 1/1 to: AK055905 from: 1 to: 3327

1079 SerGlnSerThrPheSerSerGlnAlaProAspProLysValThrSe 1095
 3 TTCACAGAGTTCGTGATTTCTACGAATGTACAGACCAAAAGATTCCAAA 52
 1095 rMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaGlyL 1112
 53 TATGTTTGAATTTATCCGTGCTTTCCGCCAACAGCAATTTTGGCGAGC 102
 1112 euLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGlyAla 1128
 103 TTGTGTTAAACAGAGCGTGTGATTTTAGACCCCTGATGCTGAGAGACG 152
 1129 PheLeuLeuHisGlyAlaHisSerAlaValHisSerLeuLeuGlySerL 1145
 153 TTGTGATTTGATTAAGAAAGTCAATATGTATCAACAATTTTACTCTCCAG 202
 1145 YHisAspThrAspPheArgTyrAlaGluAlaThrValHisAlaArgVala 1162
 203 TCACGACTCAGACCCCGGTACTCTGACCTCAGATAAAGCGTCGAGTGG 252
 1162 LagGluLeuTyrLeuProLeuLeuSerLLeaAlaArgAspThrLeuProArg 1178
 253 CCATGTTGTATCTACCTGTGATGGATTAATCATGGAACCTGACTCTCAG 302
 1179 LeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaSerLe 1195
 303 CTGTATGATTTTACAGAACTCAACAATCAACGAGAAACCAATTGTAT 352
 1195 tLeuAspSerAspThrGluGlyGluGlyAspLLeaGlyThrTLeaAsp 1212
 353 AGCCACTGATTAATGAAAGTAGAGC.....GGAAGATGATGATAAGC 396
 1212 roSerValAlaMetValAlaGlyProLeu.....Ala 1224

397 AGACCGTTGCCATGCGAATCGCAGGACATCGCTCCCTCACTAACAAG 446
 1225 ProGlySerArgAlaSerLLeuSerGlnGlyProProThrAlaSerArgAl 1241
 447 CTTGGCAGTTTCTCTCTCACGCTA.....ACAGAGGGCAGCA 484
 1241 agLYCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValL 1258
 485 ACAACACTACTTTTCAGACGAATCAAGTCGAAAGCCTTTGATCTGTCTAC 534
 1258 euThrValLeuLysAspThrGluProAlaLeuLeuGlnArgThrAlaThr 1274
 535 TTTGGGTTCTCAAAAATGCGATGAACAGTTCTACAGAAAGGTTTACA 584
 1275 AspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCy 1291
 585 GATCTCACTCTTGCAGCTAAACCGCTATTAGATCTGCTTATCTCTCG 634
 1291 gLeuAlaAlaPheGluTyrLysGlyLysAlaPheGluArgLLeaAsn 1308
 635 TGTCTCTGCTTGAATTAAGGAAAGAGTTTGAACGAATGAATA 684
 1308 euLeuThrPheLysSerLeuAspMetLysAlaArgLeuGluAla 1324
 685 GCTTGACCTTTAAGAAATCAAAAGCATGAGCAAAAGCTTGAAAGAGCT 734
 1325 tLeuGlyThrLLeuGlyAlaArgGlnGlnMetValArgArgSerArg 1340
 735 ATTCTTGGAACATAGGTGCCAGCAAGAAATGTAAGGCGCAAGCGAAG 784
 1341GluArgSerPro.....PheGlyAsnProGluAsnV 1351
 785 ACAAGCTCAGAGAAAGCCCATCTGGAAGTCCCTTGGAAATCTCAAGAAAT 834
 1351 alArgTPArgLysSerValThrHisThrLysGlnThrSerAspArgVal 1367
 835 TGAAGTGAAGGAAGATATGACTCACTGGCGTCMAAACCAAGAAAGCTT 884
 1368 AspLysThrLysAspGlnMetGlnHisGluAlaLeuValGluGlyAsnLe 1384
 885 GACAAATCAAGACAGAGATTTGAACACAGAAAGCATGATGATGAAACCT 934
 1384 uAlaThrGlnAlaSerLeuValAlaLeuAspThrLeuGluLLeaValG 1401
 935 GGCTACAGAAACCAACCTAATCATTTAGATACATTAGCATTTGTTGTC 984
 1401 tThrValMetLeuSerGluAlaArgLLeuSerValLeuGlyAlaValLeu 1417
 985 AGACCGTTCTGTACGGAATCCAAAGAGAGCATCTCTGCTGAGAGTGA 1034
 1418 LysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheLeuG 1434
 1035 AAAGTCTACTACACAGCATGCGCTTAACCAAGAGCGATTATCTACA 1084
 1434 nHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeuL 1451
 1085 ACACTGTTTGTACACAGAGAGCCTTGGTTCAAGTTTCTCGAAACCT 1134
 1451 euPheGluGluAspThrGlnLeuGlyAlaAspLeuGlyLeuArgLeuLeu 1467
 1135 TATTTGAAGAGAGACAGACAGCACTGTGATTTAAGCTCTCAGGCTTCTC 1184
 1468 ArgHisCysGlySerArgLLeuSerThrLLeuArgThrHisAlaSerAlaSe 1484
 1185 CGAACGTGTAGCAGTAGCAGGTACAAATAGCGTACACAGCGACGCGCC 1234
 1484 tLeuTyrLeuLeuMetArgGlnAsnPheGluLLeuGlyHisAsnPheAla 1501
 1235 CCTTACTACTAATAGGCAAAACCTTGAATGGAAATCACTTGGCA 1284
 1501 rGValLysMetGlnValThrMetSerLeuSerSerLeuValLysThrThr 1517

1285 GGGTAAATGCAAGTAAACATGTCATCTCTCTTGGTGGCACATCT 1334
1518 GlnAsnPheserGluGluHisLeuArgArgSerLeuLysThrIleLeuTh 1534
1335 CAGAAATTTAAAGAAATCTCTTAAGAGCTCTCTAAAGACTATATGAC 1384
1534 rTyraIaGluGluAspMetGlyLeuArgAspSerThrPheIaGluGln 1551
1385 ATATGCTGAAGAAGATCTGGAAATGAGGAAACAAACATTTCTCTGACGG 1434
1551 aIIGlaAspMetPheAsnLeuHisMetIleLeuThrAspThrValLys 1567
1435 TCCAGATCTGGTTTCAATCTCCATATGATCTTCTGATACGTGAAA 1484
1568 MetLysGluHisGlnGluAspProGluMetLeuIleAspLeuMetYrAr 1584
1485 ATGAAGAACACACAGAGATCTCTGAATGTGATGATCTAATGTACAG 1534
1584 gIleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTTPLeuG 1601
1535 AATTTGCCAAGGTTTACCAGACCTCTCCAGATCTGGACTGACCTGGTTC 1584
1601 InAsnMetIaGlyLysHisIaGluLeuGlyAsnHisIaGluAla 1617
1585 AGAACATGGCAGGCAAGCACTCAGAACGAAGCAATCATCTGAAGCTGA 1634
1618 GlnCysMetValHisAlaAlaLeuValAlaGluTyrLeuAlaLeu 1634
1635 CAGTGTCTAGTCCACTCAGACACTTGTGCTGAATATTTAGAGCTGCT 1684
1634 uGluAspArgHisLeuProValGlyCysValSerPheGlnAsnIleAs 1651
1685 GAGAGACCGGAAATATCTCTCTGGGATCTGTAAACATTTCAAGATATT 1734
1651 eSerAsnValLeuGluGluSerAlaIleSerAspSpIleLeuSerPro 1667
1735 CATCTAATGTTTGAAGAATCTGCGCTCTCAGATGATGGTATCTCA 1784
1668 AspGlnGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuVal 1684
1785 GATGAAAGAGGTATCTGCTGGAAATACTTTACTGAGTCAGAGACTGTG 1834
1684 IGIleuLeuGluGlnAlaIaGlyTyrPheThrMetGlyLysLeuTyrG 1701
1835 GGGATTCAGAACACACCTGCTCTCTCTATGCTGCGCATGTAG 1884
1701 IuAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHis 1717
1885 AAGCAGTTAATGAAGTTTACAAAGTACTTATCTTCAATTCATGAACTAAT 1934
1718 ArgAspTyrLysLysLeuAlaIaValHisGlyLysLeuGlnIuAlaPh 1734
1935 CGGGATGCAAAAGAACTATCCAAATTCATGTAACCTCAAGAGCAT 1984
1734 eThrLysIleMetHisGlnSerSerGlyTyrPGLuArgValPheGlyThr 1751
1985 CAGCAAAATGTTTCATCAGGATGT....AAGCGATCTTTGGCACCT 2028
1751 yrpheArgValGlyPheTyrGlyValHisPheGlyAspLeuAspGluGln 1767
2029 ATTTTCGCTGTTTATGAGAACCAAGTTCCGGGATTTGAGAGAACAA 2078
1768 GluPheValTyrLysGluProSerIleThrLysLeuAlaGluIleSerHis 1784
2079 GAATTTGTTTCAAGAGACCTGCAATTAACCAACTTGACAGATATCTCA 2128
1784 sarGluGlnGluPheTyrThrGluArgPheGlyAspAspValAlaGluI 1801
2129 CAGATTTGAGGATTTTACGAGAAAGATTTGAGAGAGATGTTGTAAG 2178
1801 leIleLysAspSerTyrProValAspLysSerLysLeuAspSerGlnLys 1817
2179 TAATCAAAAGACTTAATCTGTAGACAAAGTAAATTAATGATCTTAACAC 2227

1818 AlaTyrIleGlnIleThrTyrValGluProTyrPheAspThrTyrGlu 1834
2228 GCATATATTTAGATTACTATGTGGACCTACTTTGACACATATAGAT 2277
1834 uLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPhe 1851
2278 GAAGAGAGATCACTATTTCCAAATAATTTACAATCTTCGTGATTTCA 2327
1851 eupheCysThrProPheThrProAspGlyArgAlaHisGlyLysLeuPro 1867
2328 TGTACTGTACACCTTACTTGTAGATGGCCGTGCCATGGGGAATTTCA 2377
1868 GluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhePro 1884
2378 GAACAAATTCAAAGAGAGACATCTGTACATGCTCATGCTCTCTTA 2427
1884 rIleLysThrArgIleArgValCysHisArgGluGluThrValLeuThr 1901
2428 TATTAAGAACAGGCTCAATGTCACTCACTAAAGAGAGATCATTTAAC 2477
1901 roValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAla 1917
2478 CAATTTGAAGTTGCTATTGAGGACATCGAGAAAAGACAGAGGTTGCA 2527
1918 PheAlaThrGluGlnAspProProAspAlaLysMetLeuGlnMetVal 1934
2528 TTTGCAACATTCAGATCCCGCAGACCCCAAATGCTTCAGATGTACT 2577
1934 uGlnGlySerValGlyProThrValAsnGlnIleProLeuGluValAla 1951
2578 CCAGGATCTGTAGGCACACAGTGAATCAGGGCCCTTTGGAGTTGCC 2627
1951 InValPheLeuAlaGluIleProGluAspProLysLeuPheArgHis 1967
2628 AGGTTTCTGTCTGAATACCTAGTGACCCAAAGCTCTTCAGACATAT 2677
1968 AsnLysLeuArgLeuCysPheLysAspPheCysLysLysCysGluAspAl 1984
2678 AATAAACGCGACCTGCTTAAAGATTTTACTTAAAGTGTGAAGATGC 2727
1984 AleuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHis 2001
2728 CTTAAGAAAATAAGACTTAATTTGGCGGATCAAAAGAGTATCAAA 2777
2001 rGluLeuGluArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeu 2017
2778 GCGAAGTGGAGAAACTATCATCTGCTTAAAGAGGCCCTACAGCCACTG 2827
2018 LeuThrGlnArgLeuProGlnLeuMetAlaProThrProProGlyLeu 2033
2828 ATCAACAGAAAGATCCCTCAGATTATCAAGAGCATATTTGCCGTACCTG 2877
2034 ...ArgAsnSerLeuAsnArgAlaSerPheArgLysAlaAspLeu 2047
2878 CCACAGAGATCTCTTCAAGTGAATGAGCTTCGCAAAATGATCTC 2923

seq_name: gb_pat:AX172978

seq_documentation_block:

LOCUS AX172978 3253 bp DNA

DEFINITION Sequence 105 from Patent WO0142297.

ACCESSION AX172978

VERSION AX172978.1 GI:14597957

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Clasp-3 transmembrane protein

Patent: WO 0142297-A 105 14-JUN-2001;

Arbor Vita Corporation (US)

PART 03-JUL-2001

```
FEATURES                                         Location/Qualifiers
    source                                         1..3253
                                                /organism="synthetic construct"
                                                /db_xref="taxon:32630"
                                                /note="polynucleotide fragment"
BASE COUNT      1009 a      633 c      705 g      906 t
ORIGIN
alignment_scores:
    Quality: 3508.50                      Length: 952
    Ratio: 4.142                          Gaps: 10
    Percent Similarity: 88.971             Percent Identity: 72.059
alignment_block:
US-09-736-968a-2 x AX172978 ..
Align seg 1/1 to: AX172978 from: 1 to: 3253
1111 GlyLeuLeuLeuThrgLLeuLeuAlaLeuGluProGluAlaGluGlu 1127
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
4 GCACCTGTGTTCAGACGCTGCTGCATTTTACACCTCATGATCGAAGG 53
1127 yAlaPheLeuLeuEukLysLysAlaAlaIleSerAlaValAlaHisSerLeuLeuC 1144
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
54 ACTGTTGGATTGTCATAGAAAGTCAATATGCTACACAAATTACTCT 103
1144 yseGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArg 1160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
104 CCAGTCACGACGTCAGACCGCGGTACTGACCCCTCAGATTAAGGCTCGA 153
1161 ValAlaGluLeuTyrLeuProLeuLeuSerLleAlaIArgAspThrLeuPr 1177
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
154 GTGGCCATGTGTATCTACCTCGATTGATATCATGATAAGAACTGTACC 203
1177 oArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaAS 1194
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
204 TCAGCTGTATGATTTTACAAAGAACCTCAACATCAGCAGAGAACGCAATT 253
1194 erMetLeuAspSerAspThrGluGlyGluGlyAspLleAlaGlyThrIle 1210
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
254 GTATAGCCACTGATGATTAAGAAAGTAGAGAGC.....GGAAGTATGATA 297
1211 AspProSerValAlaMetAlaIleAlaGlyGlyProLeu..... 1223
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
298 AGCCAGACCGCTGCATGGCAATCGAGACATCGGTCCTCACTCACTAAC 347
1224 .AlaProGlySerArgAlaSerLleSerGlnGlyProProThrAlaSerA 1240
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
348 AAGGCTGGCAGATTTCCTCTCCTCAGCTCA.....ACGAGTGGCA 385
1240 rgaLagLysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys 1256
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
386 GGCAGACACTACTCTTTTCAGCAGAAATCAAGTCGAAAGCTTTTGATCTGT 435
1257 ValLeuTyrValLeuLysAsnThrGluProAlaLeuLeuGlnArgTrrpAl 1273
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
436 CTACTTTGGGTTCGCAAAAATGCAAGTGAAGAACTTCTACAGAAAGTGTT 485
1273 aThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrL 1290
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
486 TACAGATCTCTCACTCTTGCAGCTAAACGGCTATTAGATCTGCTTATC 535
1290 euCysLeuAlaAlaPheGluTyrLysGlyLysLysAlaPheGluArgIle 1306
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
536 TCTGTGTGTCTTGTGCTTGAAGTAAAGGGAAGGTTTGAAGCATG 585
1307 AsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArgLeuGluGlu 1323
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
586 AATGTGCTTGACCTTTAAGAAATCAAAAAGACGCAAAAGCTTGAGGA 635
1323 uAlaIleGluGlyThrLleGlyAlaArgGlnGluMetValArgArgSerA 1340
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
636 AGCTATTCTTGGAGCATAGTGGCCAGGCAAGAAATGTTACGGCGAAGCC 685
1340 rg.....GluArgSerPro.....PheGlyAsnProGlu 1349
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
686 GAGACAGAGCTCGAGAGAAAGCCCATCTGGAAGTGCCTTGGAAAGTCAAGAA 735
1350 AsnValArgTrrpArgLysSerValThrHisTrrpLysGlnThrSerAspAr 1366
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
736 AATTGGAGGTGGAGGAAAGATATGACTGCTGGGCTCAAAACACAGGAA 785
1366 gValAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyA 1383
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
786 GCTTGACAAATCAAGACAGACAGATGGAACACAGAACCTGATGATGAGA 835
1383 snLeuAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIleIle 1399
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
836 ACCTGGCTACAGAGCAACCACTATCATTTTACATTAAGATGATGTT 885
1400 ValGlnThrValIleMetLeuSerGluAlaArgLysSerValLeuGlyAlaVal 1416
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
886 GTTCAGACCGTTTCTGTAAACGAATCCAAAGAGAGCATTCCTGTGGAGT 935
1416 lleuLysValValLeuTyrSerLeuGlySerAlaGlnSerAlaLeuPheL 1433
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
936 GCTAAAGTCTACTACACACATGCGCTGTAAACCAAGTGCAGATTATC 985
1433 euGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGlu 1449
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
986 TACAACACTGTTTTCCTACACAGAGAGCGCTTGTCAAGATTCCCGAA 1035
1450 leuLeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLysAlaArgLe 1466
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1036 CTCTATTGAGAGAGAGACAGACAGAGTGTCTGATTTATGCTCAGAGCT 1085
1466 uLeuArgHisCysGlySerArgLleSerThrIleArgThrHisAlaSerA 1483
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1086 TCTCCGACAGCTGACAGTACAGATCGGTACAAATAGCGACACACCCAGTG 1135
1483 laSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPhe 1499
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1136 CCTCCCTTACTACTAATGAGGCAAACTTTGAGATTGGAGAAATCTTT 1185
1500 AlaArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyTh 1516
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1186 GCCAGGTTAAATATCAGGTACCAATGTCACTATCTCTTGGGGCCAC 1235
1236 ATCTCAGAAATTTTATGAAGAAATCTTAAGACGTTCTTAAGACTATAT 1285
1516 rThrGlnAsnPheSerGluGluHisLeuArgArgSerLeuLysThrIleL 1533
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1533 euThrTyrAlaGluGluAspMetGlyLeuArgAspSerThrPheAlaGlu 1549
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1286 TGACATATGCTGAAGAAATCGAATGAGGGAACAACTTCCGTAT 1335
1550 GlnValGlnAspLeuMetPheAsnLeuHisMetLleLeuThrAspThrVa 1566
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1336 CAGGTCAGAGATCTGTTTCAATCTCCATATGATTTCTTGTGATACGT 1385
1566 lLysMetLysGluHisGlnGluAspProGluMetLeuIleAspLeuMet 1583
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1386 GAATAATGAAGAACACACAGAGATCTGAATAATGTTGATGATCAATGT 1435
1583 yTrArgIleAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTrp 1599
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1436 ACAGAAATGGCAAGGTTTACAGACCTCTCCAGATCTCGATTGACCTGG 1485
1600 LeuGlnAsnMetAlaGlyLysHisAlaGluLeuGluGlyAsnHisAlaGluAl 1616
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1486 TTGCGAAGATCGCAGGCAAGCACTCAGACAGACATCATCTGCGAAGC 1535
1616 aaLagLncysMetValHisAlaAlaAlaLeuValAlaGluTyrLeuAlaL 1633
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1536 TGCACAGTGTCTAGTCCACTCGACAGCACTGTGTCTGAATATTTGAGCA 1585
```


OM of: US-09-736-968a-2 to: EST:* out_format: pfs
Date: Oct 5, 2002 12:46 PM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODL=framed-p2n.model -DEV=xlh
-O=/cgn2_1/USPRO.spool/US09736968/runat_02102002_145334_1322/app-query.fasta_1.2139
-DB=EST -QPM=fastap -SUFFIX=rest -GAPOP=12.000 -GAPEXT=4.500
-MINNAT=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -EGAPOP=6.000
-EGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEX=7.000 -START=1 -MATRIX=blsoms62 -TRANS=human40.cdi
-LIST=45 -DOCLLIG=200 -THR=MAX=100 -THR_MIN=0
-ALLEN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZ=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09736968.GCGL1_16698
-NCPU=6 -ICPU=3 -LONGLOG -DEV=US09736968 -WARN_TIMEOUT=30
-NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-736-968a-2
Query Length: 2047
Database: EST:*
Database sequences: 13736207
Database length: 1841457050
Search time (sec): 4657.170000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentaion
gb_hnc:AK016495	+ 4089.00	6097.36	0.0	2710	AK016495 Mus musculus adult ma
gb_hnc:BC020473	- 2212.00	3288.84	6.1e-174	1510	BC020473 Homo sapiens, clone 1
gb_hnc:AK010755	+ 2035.50	3024.09	3.4e-159	1507	AK010755 Mus musculus ES cells
gb_hnc:AK004816	+ 1819.50	2692.75	9.7e-141	2694	AK004816 Mus musculus adult ma
gb_hnc:BC018599	+ 1574.50	2334.25	9.0e-121	1311	BC018599 Mus musculus, clone 1
gb_est2:BT770146	+ 1364.00	2024.71	1.6e-103	796	BT770146 6030533322F1 NIH.MGC_12
gb_est2:BT820302	+ 1304.50	1933.54	1.9e-98	927	BT820302 603030684F1 NIH.MGC_11
gb_est2:BT830310	+ 1284.00	1905.12	7.2e-94	770	BT830310 603073048F1 NIH.MGC_11
gb_est2:BE551136	+ 1256.00	1863.55	1.7e-94	805	BE551136 601278135F1 NIH.MGC_39
gb_est2:BT854143	+ 1237.00	1863.45	7.0e-93	842	BT854143 603381452F1 NIH.CGAP_M
gb_est2:BT078974	+ 1213.00	1799.49	6.3e-91	777	BT078974 602873230F1 NCI.CGAP_M
gb_est2:BT181383	+ 1211.00	1795.62	9.0e-91	769	BT181383 UNL-P-FN-a8-08-0-UNL
gb_est2:BG678867	+ 1193.00	1767.26	3.4e-89	857	BG678867 60264767F1 NCI.CGAP_S
gb_est2:BT191959	+ 1178.00	1749.92	5.3e-88	778	BT191959 602922704F1 NIH.CGAP_M
gb_est2:AW742035	+ 1151.00	1706.93	7.9e-86	692	AW742035 up52604.Y1 Soares-mous
gb_est2:AJ397707	+ 1127.00	1669.17	1.0e-83	796	AJ397707 AJ397707 dkfz426.Gallu
gb_est2:BT825534	+ 1118.00	1656.50	5.1e-83	745	BT825534 603035242F1 NIH.MGC_11
gb_est2:BE857715	+ 1116.00	1655.10	6.0e-83	654	BE857715 7946603.X1 NCI.CGAP_P
gb_est2:AT653716	- 1108.00	1647.98	2.9e-82	662	AT653716 wb3607.X1 NCI.CGAP_G
gb_hnc:AK018051	+ 1104.50	1625.81	2.6e-81	1712	AK018051 Mus musculus adult ma
gb_est2:BT181397	- 1090.00	1613.37	1.3e-80	815	BT181397 603076025F1 NIH.MGC_11
gb_est2:BT110605	+ 1079.00	1598.32	8.6e-80	714	BT110605 60286211F1 NCI.CGAP_M
gb_est2:BE647357	+ 1065.50	1578.59	1.1e-78	696	BE647357 UI-M-BHL-amg-d-03-0-UI
gb_est2:BG691673	+ 1064.50	1577.42	1.3e-78	678	BG691673 340944.BARC.5BOV.Bos
gb_est2:BM473590	- 1057.00	1559.97	1.2e-77	1112	BM473590 AGECOCURT.646302.NIH
gb_est2:AT582283	- 1045.00	1549.81	5.1e-77	644	AT582283 tg65109.X1 NCI.CGAP_L
gb_est2:AM516592	+ 1043.00	1546.45	6.8e-77	612	AM516592 xg01h03.X1 Soares.NHCG
gb_hnc:AK016777	+ 1043.00	1538.74	1.8e-76	1132	AK016777 Mus musculus adult ma
gb_est2:BM023626	+ 1011.00	1498.95	3.0e-74	588	BM023626 ieb2912.Y1 Melton.Norm
gb_hnc:BC018075	+ 1008.00	1479.92	3.5e-73	1873	BC018075 Homo sapiens, clone 1
gb_hnc:BC018076	+ 1008.00	1479.92	3.5e-73	1873	BC018076 Homo sapiens, clone 1
gb_est2:BF980727	+ 1003.50	1481.84	2.7e-73	938	BF980727 602303869F1 NIH.MGC_8
gb_est2:BG914864	+ 999.00	1479.55	3.6e-73	657	BG914864 602813690F1 NIH.CGAP
gb_est2:BG247108	+ 986.00	1455.68	7.8e-72	931	BG247108 602359776F1 NCI.CGAP
gb_est2:AM027279	+ 979.00	1450.44	1.5e-71	612	AM027279 wv66003.X1 Soares.thym
gb_est2:BM126124	- 977.00	1448.02	2.1e-71	584	BM126124 if05h06.X1 Melton.Norm
gb_est2:BG912304	+ 972.50	1438.97	8.1e-71	789	BG912304 602806839F1 NCI.CGAP_H
gb_est2:AW186286	+ 970.00	1437.42	6.1e-71	702	AW186286 xg62609.X1 NCI.CGAP_U
gb_est2:BF568737	+ 961.00	1432.84	4.1e-70	547	BF568737 602184333F1 NIH.MGC_4
gb_est2:BB659390	+ 961.00	1423.60	4.8e-70	604	BB659390 BB659390 RIKEN full-16

gb_est2:BF569035 - 961.00 1421.59 6.2e-70 709 | BF569035 602184383F1 NIH.MGC
gb_est2:BF115265 + 960.00 1422.83 5.3e-70 570 | BF115265 7n80f03.X1 NCI.CGAP
gb_est2:AW369859 - 947.00 1404.21 5.7e-69 531 | AW369859 QV0-BT0229-251099-0
gb_est2:BT737331 + 946.00 1398.20 1.2e-68 761 | BT737331 603357985F1 NIH.MGC
gb_est2:BM126416 + 943.00 1397.17 1.4e-68 577 | BM126416 if05h06.Y1 Melton.N

seq_name: gb_hnc:AK016495

seq_documentation_block:

LOCUS AK016495 2710 bp mRNA linear HTC 19-JAN-2002

DEFINITION Mus musculus adult male testis cDNA, RIKEN full-length enriched

library, clone:4931431C02:homolog to KIAA1395 PROTEIN (FRAGMENT),

full insert sequence.

ACCESSION AK016495.1 GI:12855259

KEYWORDS HT: CAP trapper.

SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA,

clone.lib:RIKEN full-length enriched mouse cDNA library

clone:4931431C02.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (sites)

Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,

Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M., and Hayashizaki,Y.

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes

Genome Res. 10 (10), 1617-1630 (2000)

3 (sites)

Shibata,K., Itoh,M., Aizawa,K., Nagakura,S., Sasaki,N., Carninci,P.,

Konno,H., Akiyama,U., Nishi,K., Kitamura,T., Tashiro,H., Itoh,M.,

Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishimura,T., Harada,A.,

Yamamoto,N., Matsunoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,

Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M.,

Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A., and Hayashizaki,Y.

RIKEN integrated sequence analysis (RISA) system-184-format

sequencing pipeline with 384 multicapillary sequencer

Genome Res. 10 (11), 1757-1771 (2000)

4 (sites)

The RIKEN Genome Exploration Research Group Phase II Team and the

FANTOM Consortium.

Functional annotation of a full-length mouse cDNA collection

Nature 409, 685-690 (2001)

5 (bases 1 to 2710)

Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,

Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bull,C.,

Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Haraagaki,T.,

Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F.,

Hume,D., Imotani,K., Ishii,Y., Itoh,M., Iwama,M., Kasuawa,T.,

Kato,H., Kawai,D., Kojima,Y., Konno,H., Kouda,M., Komura,S.,

Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,

Nunataki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,

Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,

Schmidt,L., Shibata,K., Shibata,Y., Shingawa,A., Shiraki,T.,

Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,

Tanaka,T., Tejima,Y., Toyota,T., Yamamura,T., Yamane,K.,

Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M., and

Hayashizaki,Y.

Direct Submission

Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of

Physical and Chemical Research (RIKEN), Laboratory for Genome

Exploration Research Group, RIKEN Genomic Sciences Center (GSC),


```

|||||
952 TACTGAAGACATTGTGAAAGTCCCTCAAGACATCTGACCTACCGCAG 1001
1537 LUGLAspMetGlyLeuArgSerThrPheAlaGluInValGlnAsp 1553
1002 AGGAGGACATAGGGCTGAGGAGACAGCAGCTCGCTGAGCAGGTCAGCAG 1051
1554 LeuMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysG 1570
1052 CTCATGTTCAACCTGCATCATGTATCTGCAGACACGCTGAATGATGAGGA 1101
1570 uHisGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAla 1587
1102 ACACGAGGAGGAGGACCTGAGATGCTCATGAGCTCATGTACAAATTGCTC 1151
1587 rGlyTyrGlnGlySerProAspLeuArgLeuThrTyrLeuGlnAsnMet 1603
1152 GGGGATACCAAGGCTCTCCAGACCTGGGCTGACATGGCTGCAGAAACATG 1201
1604 AlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaGlnCysMe 1620
1202 GGTGCAAAACATGCAAGCTGGGCAATCATGCAGAGGCTGCCAGTGCAT 1251
1620 tValHisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuGluAspG 1637
1252 GGTCCACGGCGCGCTGGCTGGCCCAATACCTTGGCCCTCTCTCGAGACA 1301
1637 LrArgHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsn 1653
1302 GTCCGACACCTGCTGGGCTGTGTCTCTTCCAGAAATGTCATCAAC 1351
1654 ValLeuGluGluSerAlaIleSerAspAlaIleLeuSerProAspGlu 1670
1352 GTGCTGGAAGATGTCGCAATCTGTATGACATCTGTCCGAGATGAGGA 1401
1670 uGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeu 1687
1402 GGGCTCTCTCTAGSAAACTTCAACAACATGGGTCTGGTGGGGCTGC 1451
1687 eUGLInAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaVal 1703
1452 TGGAGAGGCGGCGGCTACTTCACCATGGGTGTGTATGAAACGGG 1501
1704 AsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHisArgAsp 1720
1502 AACGAGCTTACAAACCTTATCCCATCTTGAAGCCCAAGACGACTA 1551
1720 rLysLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLys 1737
1552 CAAGAAGCTGGCTGGTGCAGGGAACCTGCAGAGGCTTCCACCAAGA 1601
1737 lMetHisGlnSerSerGlyTyrGluArgValPheGlyThrTyrPheArg 1753
1602 TTATGCACACGAGCTCTGCTGGAGCGGTGTATGGACATATTTCCGA 1651
1754 ValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPhe 1770
1652 GTGGGCTTATGACACAGATTGTGATCCTGGATGAACAAGACTTTGT 1701
1770 lTyrLysGluProSerIleThrLysLeuAlaGluIleSerHisArgLeu 1787
1702 GTACAAAGGACCGTCAATCAGCAACCTTGCAGAGATCTCACACCGGCT 1751
1787 lGluGluPheTyrThrGluArgPheGlyAspAspValAlaGluIleLys 1803
1752 AGGACTTCTATACGGAAGGTTCCGGGATGATGTGTAGAGATCATCAA 1801
1804 AspSerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyr 1820
1802 GATTTCACCCAGTGCACAGTGCAGGCTGAGCCACGAAGGCGGTACAT 1851
1820 eGlnIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAsp 1837
|||||

```

```

1852 ACAGATTAACCTATGTGGACCAATTTGACACTTATGAGCTCAAGATC 1901
1837 rGValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCys 1853
1902 GGGTACCTACTTCGATTCGGAACATATGGCTGGGGCTTCTCTTCTTGC 1951
1854 ThrProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnH 1870
1952 ACACCTTCACACACAGATGAGAGCTGGCCGACGAGAGATGGCCGAACAGA 2001
1870 sLysArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysT 1887
2002 CAAGCAGAGAGGCTGCTGAGACGAGACATGCTTTCCTTCTCATCAGA 2051
1887 hrArgIleArgValCysHisArgGluGluThrValLeuThrProValGlu 1903
2052 CACGATCCGAGTGTGCCACCGTGAAGAGACAGTGTGACACCGAGTGAG 2101
1904 ValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAla 1920
2102 GTGGCCATTTGAGGACATGCAGAAAGACCCGGGAGCTGGCTTGGCCAC 2151
1920 rGluGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnGly 1937
2152 CGAGCAGACCTCTCAGATGCAAGATGCTGAGATGATGCTCCAGGGTT 2201
1937 eValGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPhe 1953
2202 CTGGGGAGCCACAGTGTGAACAGGTCCTTGGAAAGTGGCCAGCTGTT 2251
1954 LeuAlaGluIleProGluAspProLysLeuPheArgHisHisAsnLys 1970
2252 TTGTGAGAGATCCAGAAAGATCCCAAGCTTCCGACATCAACAACACT 2301
1970 uArgLeuCysPheLysAspPheCysLysLysCysGlyAspAlaLeuArg 1987
2302 CCGGCTGCTTCAAGATTTTCTCAAAAGTGGAGAGATGACATGGAA 2351
1987 yAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeu 2003
2352 AGAACAGAGCCCTGATTGGCCAGACCAAGAGATACCAACGGGAGCTG 2401
2004 GluArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThr 2020
2402 GAGGCTCATTATAGCCCGCTGGGAGGCTTGCACGCTCTGCTTACCA 2451
2020 nArgLeuProGlnLeuMetAlaProThrProProGlyLeuArgAsnSer 2037
2452 ACGTCTGCCCAAGCTGCTGGCACCACCACTTCCACCAGCCTCAGAGCTCA 2501
2037 eAsnArgAlaSerPheArgLysAlaAspLeu 2047
2502 TGAACAGATCAAGTTTCAGGAAGGCTGACCTC 2533
seq_name: gb_hc:BC020473
seq_documentation_block:
LOCUS BC020473 1510 bp mRNA linear HTC 03-JAN-2002
DEFINITION Homo sapiens, clone IMAGE:3853958, mRNA.
ACCESSION BC020473
VERSION BC020473.1 GI:18042973
KEYWORDS HTC.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1510)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-JAN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

```

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villation@bcm.tmc.edu
Villation, D.K., Luna, R.A., Hale, S.M., Huiyk, S., Lu, X., Garcia,
A.M., Hollaway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gunaratne, P., Yoon, V., Kowls, C., Martin, R.,
Lawrence, S., Richards, S., Gibbs, R.A.

FEATURES
SOURCE
1. 1510
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3853958"
/tissue_type="Colon, adenocarcinoma"
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
BASE COUNT 287 a 431 c 456 g 336 t
ORIGIN

alignment_scores:
Quality: 2212.00 Length: 425
Ratio: 5.229 Gaps: 0
Percent Similarity: 99.529 Percent Identity: 99.294

alignment_block:

US-09-736-968a-2 x BC020473/rev ..

Align seg 1/1 to reverse of: BC020473 from: 1 to: 1510

```
1623 AAlaAlaLeuValAlaGluTyrLeuAlaLeuLeuGluAspGlnArgHt 1639
|||||
1510 GCGGCGCGCTCGTGGCTGATACCTCGCCCTGCTGAGAGACACCGCA 1461
|||||
1639 GTGCTCGGGAAGCATTCACGTGAGCGCTGAGGCTTCTGGAACA 1311
|||||
1460 CCTGCCCGTGGGCTGCTTCTCCAGAACATCTCATCATGCTGCTAG 1411
|||||
1656 TugLusSerAlaIleSerAspAlaLeuSerProAspGluGlyPhe 1672
|||||
1410 AGGAGTCCGCATCTCCGAGACATCTGCGCCGACGAGAGGCTTC 1361
|||||
1673 CysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuGluGlu 1689
|||||
1360 TGCTCCGGAAGCATTCACGTGAGCGCTGAGGCTTCTGGAACA 1311
|||||
1689 nAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluV 1706
|||||
1310 GGCAGCGGCTACTTACCATGGGCGGCTCTACGAGCGGTGATGAGG 1261
|||||
1706 aTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLys 1722
|||||
1260 TCTACAGAAACATCCCATCTCTGAGACCCACCGTACTACAAAG 1211
|||||
1723 LeuAlaAlaValHisGlyLysLeuGluGluAlaPheThrLysIleMetH 1739
|||||
1210 CTGGCGCGGTGACGCAAACTGCGAGGCGCTTACCAAGATCATGCA 1161
|||||
1739 sGlnSerSerGlyTyrGluArgValAlaPheGlyThrTyrPheArgValGlyP 1756
|||||
```

```
1160 CCAAGTTCGCGCTGGAGCGCGTGTTCGGAGCATTTTCCGCGTGGCT 1111
|||||
1756 heTyrGlyAlaHisPheGlyAspLeuAspGluGluPheValTyrLys 1772
|||||
1110 TCTAGCGCGCCACTTGGTTGGCTGGATGACGAGGATTGTGTACAG 1061
|||||
1773 GluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGluGluP 1789
|||||
1060 GAGCCATCATCATCAGAGAGTGGGAGAGATCTCACACCGCTGGAGGAT 1011
|||||
1789 eTyrThrGluArgPheGlyLysAspValAlaGluIleIleLysAspSer 1806
|||||
1010 CTACACGAGAGATTGGCGACGCGTGTGAGATTATCAAGACTCTA 961
|||||
1806 YrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIle 1822
|||||
960 ACCCTGTGGACAGTCCAGCTTGTACTACAAAAGGCTCATCCAGATC 911
|||||
1823 ThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArgVal 1839
|||||
910 ACGTATGTGGAACCGTACTTGTATCTACAGCTCAAGAGCCGGGTGAC 861
|||||
1839 rTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProp 1856
|||||
860 CTACTTTGACCGCAACTATGAGGCTTCCACATTCCTGTTGACAGCGCT 811
|||||
1856 heThrProAspGlyArgAlaHisGlyLysLeuProGluGlnHisLysArg 1872
|||||
810 TCACGCGCGATGGCGCGCACGCGGAGGAGCTGCCGCAACAAAGCGT 761
|||||
1873 LysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrArg 1889
|||||
760 AAGAGCGTGTCTCAGCACCGACCGACCTTCCCTCATCAAGACTGCGAT 711
|||||
1889 eArgValCysHisArgGluGluThrValLeuThrProValGluValAla 1906
|||||
710 CCGTGTGTGCCACCGGAGAGAGAGGCTGCTGACGCCGCTGAGAGTGCCA 661
|||||
1906 IeGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluGln 1922
|||||
660 TCGAGGACATCGACAGAGACACGCGAGCTTGCTTCCACCGAGCAG 611
|||||
1923 AspProProAspAlaLysMetLeuGlnMetValLeuGlnIleGlySerVal 1939
|||||
610 GACCACACAGATGCTAAGATCTACAGATGCTTACGAGGCTCTGTAGG 561
|||||
1939 YrProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeuAla 1956
|||||
560 GCCACCGCTGAACAGAGGTCCCTCGAGAGTGGCCACAGTGTTTTACAG 511
|||||
1956 LuIleProGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeu 1972
|||||
510 AGATCCCGGAGACCCCAAGCTTTCGCGCATCTCAACAAATTGCGGCTC 461
|||||
1973 CysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLysAsn 1989
|||||
460 TGTCTCAAGACTTCTCAGAAATGTGAGATGCGCTGCGGAAATTA 411
|||||
1989 sAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGluArg 2006
|||||
410 GGCCTGATTGGCGCGGACGAGAGAGTACCAACGAGAGCTGAGCGCA 361
|||||
2006 sNTrYrCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeu 2022
|||||
360 ACTACTGCGCGCTCGGAGAGCTCTGAGGCCCTTACCAAGCGCGCTG 311
|||||
2023 ProGlnLeuMetAlaProThrProProGlyLeuArgAsnSerLeuAsnAr 2039
|||||
310 CCCCAGCTGATGSCACCCACCCACCGGCTCAGGAATCTTTGAACAG 261
|||||
2039 gAlaSerPheArgLysAlaAspLeu 2047
|||||
260 AGCAAGTTTCGAAAGCGACGACTC 236
```


253 TCTACATGTCAGCTGCCAGCCCCGGCCAGGACAGCAAGCCTTGAGAGCGCT 302
654 VALGlyPheThrTrpIleProLeuLeuGlnHisGlyArgLeuArgTrpG 670
303 GTGGGCTTCACCTGATCCCTCTCTGTACAAACATGCGCGCTGAGAGCATGG 352
670 YProPheYLeuProValSerValAspGlnProProSerYLeuYSerV 687
353 TCCCTTGCCCTGCGCGGTGCTGGAGACACCGCTCCACCCAGCTACTCGG 402
687 aLeuThrProAspValAlaLeuProGlyMetArgTrpValAspGlyHis 703
403 TCCGTAACCCAGATGTAGCGCTGCTGGCATGCGCTGGGTGAGAGCGCAC 452
704 LysGlyValPheSerValGlyLeuThrAlaValSerSerValHisProG 720
453 AAGGCGGTTCAGTGTGAGCTCACCGCCGTGTGTCCTGTCAGCCGCA 502
720 nAsProTrpLeuAspLysPhePheThrLeuValHisValLeuGluG 737
503 GACACCCCACTTGATAGTTCTTCACCGCTGCTACATCTTCTAGAGGAG 552
737 YAlaPheProPheArgLeuLysAspThrValLeuSerGlyLysVal 753
553 GGATCTTCCGTTCCGACTCAAGGAGACAGTGTAGCTAGGCACTATG 602
754 GluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProG 770
603 GAGCGAGGATGGCGAGCGAGCTGGACCCCTGGCCCTTGACGCCGCA 652
770 uProLeuValAlaPheSerHisValLeuAspLysLeuValArgLeuV 787
653 GCCCTAGTAGCCTTTCCACCTGCTGCTGACAGAGTTTCCGCTGG 702
787 aLLeaArgProProIleLeuSerGlyGlnIleValAsnLeuGlyArgGly 803
703 TTGTGCGGCGACCCATCATTTGTGTGGCAGATGTGAACCTGGTGCAGAGG 752
804 aAPheGluAlaMetAlaHisValAlaSerLeuValHisArgSerLeuG 820
753 GCGTTTGAAGCTATGCTATGTACACGCTGTGCACGCGAAGCTTGA 802
820 uAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaLeuArgValH 837
803 GGCTGTCCAGAGCTCCGCTGGCCACTGGCCACTGCTGGCTCCATGTGCC 852
837 aSTyrAlaPheArgLeuProGlyThrGluProSerLeuProAspGlyAla 853
853 ACTAGCGCCTTTCGCTCCTGCTGTGTGACTTCAAGCTCCAGCGCAAGCC 902
854 ProProValThrValGlnAlaAlaThrLeuAlaArgGlySerGlyArgPr 870
903 CCTCCGACAGCTGTGAGGCTCCCACTGGCCCTGGCTGTGTCGCGCC 952
870 oAlaSerLeuYrLeuAlaArgSerLysSerLieserSerSerAsnProA 887
953 CGCAGACCTGTACCTGGCAGCGCTTAAGAGCATCAGACAGACACACCTG 1002
887 sPLeuAlaValAlaProGlySerValAspAspGlyValSerArgIleLeu 903
1003 ACCTGGCTGTGGTCCCTGTGCTGTGGAGAGATGTGTCGCGCATCTTA 1052
904 aLaser..... 905
1053 GCCAGCAGGCCATGACCGCAACTGTAGCGGGCTCTTACTCTGA 1102
906L 906
1103 GGCGTCATCCTGGGCGCTCCAGCCCCCTCAGTCGAGACACACTGTGAGAGA 1152
906 yslLeuLeuHisGluGluLeuAlaLeuGlnTrpValValSerSerSerAla 922
1153 AGCTCTGTCAGCGAGAGTGGCTGTGCACTGTGGTGTGTGAGCGGAGTGGCG 1202

923 ValArgGluAlaIleLeuGlnHisAlaTrpPhePheGlnLeuMetVa 939
1203 GTTCGTGAGCTAGTCTCGACAGATGCGTGTCTTCTTCCACTATGAT 1252
939 LysSerMetAlaLeuHisLeuLeuLeuGlnGlyAlaArgLeuAspThrProA 956
1253 AAAAGCATGAGAGCTCATCTGCTGTGGCCAGCGAGTGTGACACTGCC 1302
956 rGlyLeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuVal 972
1303 GCAACGTGGCTTCCCTGGGCGCTGCTGTGATGACATTGCCCCCTGGTG 1352
973 GlySerValGlyLeuGluValIleThrArgValHisLys 985
1353 GCTTCGTGGCGCTGGAAGTCATCACCAGATGCCATTAAG 1391
seq_name: gb_hlc:AK004816
seq_documentation_block: 2694 bp mRNA linear HTC 19-JAN-2002
LOCUS AK004816
DEFINITION Mus musculus adult male lung cDNA, RIKEN full-length enriched library, clone:1200017A24;homolog to BA165P24.1.1 (NOVEL PROTEIN SIMILAR TO DROSOPHILA CG6630 AND CG11376, KIAA1058, RAT TRG (ISOFORM 1)) (FRAGMENT), full insert sequence.
ACCESSION AK004816 GI:12836282
VERSION AK004816
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (strain: C57BL/6J) adult male lung cDNA to mRNA, clone:11b-RIKEN full-length enriched mouse cDNA library
clone:1200017A24.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (sites)
AUTHORS Carninci,P. and Hayashizaki,Y
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2 (sites)
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3 (sites)
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kasaiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Tomoda,Y., Ishikawa,T., Kira,A. and Hayashizaki,Y.
TITLE Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichipillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4 (sites)
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5 (bases 1 to 2694)
AUTHORS Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Balderelli,R., Bono,H., Brownstein,M., Bulic,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hirooka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,


```

1818 laTyrIleGlnIleThrTyValGluProTyrPheAspThrTyGluLeu 1834
|||||
845 CCTACATTGACATCATTTGTGAGCCTTACTTGTATGAAATATGAGATG 894
|||||
1835 LysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLe 1851
|||||
895 AAAGACCGGGTGACCTTCTTGAGAAAGAAATTCAACTCCGAGGTTTCAT 944
|||||
1851 uPheCysThrProPheThrProAspGlyArgAlaHisGlyGluLeuPro 1868
|||||
945 GTACACCCACCCCTTCACCCCTGAGAGGAGACCCCGGGGGAGCTTCATG 994
|||||
1868 LngHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProTyr 1884
|||||
995 AGCAACACCCGAGAAACACCGTCTCCACCATCATGACGCTTCCCTTAC 1044
|||||
1885 IleLysThrArgGlyLeuArgValCysHisArgGluGluThrValLeuThrPr 1901
|||||
1045 ATCAAGACCGAGATCCGAGTCCAGCCAGAAAGAGAGTTCCGTTTGACTCC 1094
|||||
1901 oValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAlaP 1918
|||||
1095 GATTGAGGTTCCCATGTGAAGATATGAGAAAGAACACCTCGACTTACCG 1144
|||||
1918 heaLarThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLeu 1934
|||||
1145 TGCGCCACTCACAGAGAGCCCTGATCAAAAGATGTCGAATGCTACTG 1194
|||||
1935 GlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAlaG 1951
|||||
1195 CAGGGCTCTGAGAGCCACTGTAAATCAGGACCTGGAGGGGCCCA 1244
|||||
1951 nValPheLeuAlaGluIleProGluAspProLysLeuPheArgHisHis 1968
|||||
1245 AGTGTCTTGCTGAATTCACAGCTGACCCAAAGCTCTACGGAATCACA 1294
|||||
1968 snLysLeuArgLeuCysPheLysAspPheCysLysLysCysGluAspAla 1984
|||||
1295 ACAAGCTGAGGTGTGCTTCAAGAGATTCAATGATGATCGAGGAGGCC 1344
|||||
1985 LeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisAr 2001
|||||
1345 GTGAGAAAGAAACAGCGCATCTATCAGCGCAGAGCGGGGATGACGCA 1394
|||||
2001 gGluLeuGlnArgAsnTyrCysArgLeuArgGluAlaLeuGlnProLeu 2018
|||||
1395 GGAGCTGAGAAAGAACTACAACAAGCTGAGAGACGCTCAGGCCCATGA 1444
|||||
2018 euThrGlnArgLeuProGlnLeuMetAlaPro.....ThrProPro 2031
|||||
1445 TTGAGCGGAAATCCAGAGCTCTACAAAGCCCATATTCTCAGAGTTGACAGT 1494
|||||
2032 GlyLeuArgAsnSerLeuAsnArgAlaSerPheArgLysAlaasp 2046
|||||
1495 CAGAAGAGGAGCACTTCTCCACAGATCTAGTTTCAAGAAATGTGA 1539
|||||
seq_name: gb_hlc:BC018599
seq_documentation_block:
LOCUS BC018599 1311 bp mRNA linear HTC 03-DEC-2001
DEFINITION Mus musculus, clone IMAGE:3596745, mRNA.
ACCESSION BC018599
VERSION BC018599.1 GI:17235789
KEYWORDS HTC.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer

```

```

REMARK
COMMENT
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: gcgabs-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villaloboscm.tmc.edu.
Villalobos, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 16 Row: e Column: 2
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: no 5' EST match.
location/Qualifiers
1..1311
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3596745"
/russue_type="Mammary tumor, C3(1)-Tag model. Infiltrating
ductal carcinoma. 5 month old virgin mouse."
/clone_1id="NCI_CGAP_Mam6"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 363 a 309 c 333 g 306 t
ORIGIN
alignment_scores:
Quality: 1574.50 Length: 440
Ratio: 4.048 Gaps: 5
Percent Similarity: 88.409 Percent Identity: 71.364
alignment_block:
US-09-736-968a-2 x BC018599 ..
Align seg 1/1 to: BC018599 from: 1 to: 1311
1197 AspSerAspThrGlnGluGlyLysPheAlaGlyThrIleAsnProse 1213
|||||
16 GACTATGACAGTGAAGT.....GGGACATGATTAAGCCAGAC 53
|||||
1213 rValAlaMetAlaIleAlaGlyGlyProLeu.....AlaProG 1226
|||||
54 GGTTCGATGCGCAATTGCGAGAAACATCCGCTCCCTCAGCTGACAAACCTG 103
|||||
1226 lYSerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGly 1242
|||||
104 GGAGTTTCTCCTCAGCTCA.....ACGAGTGGCGCGCAGCAC 141
|||||
1243 CysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValLeuTr 1259
|||||
142 ACCACCTCTCAGCAGAGTCAAGCAGAGGCCCTTTGATCTGCTGCTTGG 191
|||||
1259 pValLeuLysAsnThrGluProAlaLeuLeuGlnArgThrPalaThrAspL 1276
|||||
192 GGTTCGAAAGATGCAAGATGAACGTTCTGCAAGAAATGGTTTACAGATC 241
|||||
1276 euThrLeuProGlnLeuGlnArgLeuLeuAspLeuLeuTyrLeuCysLeu 1292
|||||
242 TCTCAGTCTCCAGCTGAACCGGTTGTAAGATCTGCTTACCTGCTGTGTA 291
|||||
1293 AlaAlaPheGluTyrLysGlyLysAlaPheGluAlaSerLe 1309

```

```

292 TCTGCTTTGAGTACAAAGGAAAAAGATTGTAACGATGAATGATT 341
1309 uThrPheTyrLysSerLeuSpmMetLysAlaArgLeuGluValIleL 1326
342 AACCTTTAGAAATCAAAAGACATGAGAGCCAAAGCTTGAGAGACCATTC 391
1326 euGIYThrIleGlyAlaArgGluMetValArgArgSerArg..... 1340
392 TGGGGGACATCGGTGCCAGGAGAAATGGTACGGCGGAGCGAGGCGAG 441
1341 ...GluArgSerPro.....PheGlyAsnProGluAsnValArg 1352
442 CTCGAGAGAGAGCCCATCTGGAAGCCCTTTGGAGAGCCAAAGAAACCTGCG 491
1352 gTTPArgLysSerValThrHisTrpLysGlnThrSerAspArgValAspL 1369
492 GTGGAGAAAGACATGACTGCTGCGTCAAGTCACTAGAGAAACCTGATA 541
1369 ySThrLysAspGluMetGluHisGluAlaLeuValGluGluysnLeuAla 1385
542 AGTCAAGACACAGAGATGACAGACGAGCTGATTTGATGAAACCTGGCT 591
1386 ThrGluAlaSerLeuValValLeuAspThrLeuGluIleIleValGlnTh 1402
592 ACTGAAGAAACCTCATCTCATCTTACAGACGCTAGAGATCATGCTCCAGAC 641
1402 rValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLeuLysV 1419
642 TGTGCTGTGAACGGAATCCAAAGAGAGTTCGTGGTGGTGTAAAG 691
1419 aValLeuLysSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHis 1435
692 TGTGCTGTGAAGCATGGCTGCCAACCAAGTGCAGTCTATCTGCAGCAC 741
1436 GlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeuLeuPh 1452
742 TGTGTCGCCACGAGAGAGCCCTGGTCTCAAAAGTTTCCGTGACTCTGTT 791
1452 eGluGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgH 1469
792 TGGGGAAGAGACAGACAGCTGCGGATCTGCTCCCGCTTCGCCGTC 841
1469 iScylGlySerArgLysSerThrIleArgThrHisAlaSerLysLeuLeu 1485
842 ATTTGAGTACAGCATCATGATACATCCGGTCGACCTAGTCCCTCCTT 891
1486 TyrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgVa 1502
892 TACCTGCTCATGAGCGAAGACTTGGATCGGGAATAACTTTGCCAGAGT 941
1502 LysMetGlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAla 1519
942 GAAATAGCAAGTGACGATGCTCACTGCTCTTGGGGGTACGTCACAGA 991
1519 snPheSerGluGluHisLeuArgArgSerLeuLysThrIleLeuThrTyr 1535
992 ATTTTATATGAAGATTCTTAAGACGCTCTATAAAACATAATTTGACATAT 1041
1536 AlaGluGluAspMetGlyLeuValArgAspSerThrPheAlaGluGlnValG1 1552
1042 GCGAAGAAAGATCTAGAAATTGAGGGAACACATTTCCGTATCAGGTCCA 1091
1552 nAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetL 1569
1092 AGATTGTGCTTCATATCCATATGATCTTCTCTGACACTGTTAAGATGA 1141
1569 yGluHisGlnGluAspProGluMetLeuIleAspLeuMetLysArgLys 1585
1142 AGGAGCATAGGAAGACCCAGAAATGCTATTGATCTCATGTAAGAAATC 1191
1586 AlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAs 1602

```

```

1192 GCCAAGGCTACAGACGTCTCCAGACCTGCGCTGACTGTTGCAGAA 1241
1602 nMetAlaGlyLysHisAlaGluLeuGluGlyAsnHisAlaGluAlaGlnC 1619
1242 CATGCTGGCGAAGCATCTGTAACGAAACATTCATGCCGAGCTGTCTCAGT 1291
1619 ySMetValHisAlaAlaAla 1625
1292 GCGTGTCTCACTCCGAGACC 1311

```

seq_name: gb_est2:BI770146

seq_documentation_block: 796 bp mRNA linear EST 25-SEP-2001
 LOCUS BI770146
 DEFINITION 603053352F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5203063 5',
 mRNA sequence.
 ACCESSION BI770146
 VERSION BI770146.1 GI:15761724
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 796)
 NIH-MGC http://imgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLNL1508 row: p column: 08
 High quality sequence stop: 793.

FEATURES

source

```

1..796
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5203063"
/clone_lib="NIH_MGC_122"
/lab_host="DH10B"
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;
site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
```

BASE COUNT 188 a 228 c 231 g 148 t 1 others
 ORIGIN

alignment_scores:

Quality: 1364.00 Length: 264
 Ratio: 5.226 Gaps: 2
 Percent Similarity: 98.864 Percent Identity: 98.864

alignment_block:

US-09-736-968A-2 x BI770146 ..

Align seg 1/1 to: BI770146 from: 1 to: 796

```

1755 glyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValTy 1771
|||||
1 GCGCTTTCAGCGCGCCCACTTGGTACCTGGATGAGCAGAGATTGTGTGA 50

```

```

1771 rlysgluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGlu 1788
|||||
51 CAAGGAGCATCGATCAGACAGTGCAGATCTCACACCGCTGGAGG 100
|||||
1788 lueherythrGluArgPheGlyAspAspValValGluIleLeuLys 1804
|||||
101 AGTTCTACACGGAGAGATTGGCGAGCAGCTGTTGATTTCAAAAGAC 150
|||||
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG1 1821
|||
151 TCFAACCCGTGTGACAAAGTCAAGCTTGACTCAAAAAGCCTACATCA 200
|||||
1821 nilethrrValGluProTyrPheAspThrTyrGluLeuLysAspArgV 1838
|||||
201 GATCAGCTATGTGGAACCGTACTTGTATCTACAGAGCTCAAGACCGGG 250
|||||
1838 alThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheGlyThr 1854
|||||
251 TGACCTACTTGTACCGCAACTATGGGCTTCGACATCTCTGTTCTGCACG 300
|||||
1855 ProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisLy 1871
|||||
301 CGGTCACCGCCGGATGGCGCGACACGCGGAGCTGCCGACCAACACA 350
|||||
1871 sarGlyThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrA 1888
|||||
351 GCGTAGAGAGCTGCTCAGCACCGACCGCTTCCCTTACATCAAGACTC 400
|||||
1888 rtleArgValCysHisArgGluGluThrValLeuThrProValGluVal 1904
|||||
401 GCATCCGTGTGTGCCACCGGAGAGACGCTGTACGCCACTGGAGGtG 450
|||||
1905 AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrG1 1921
|||||
451 GCCATCGAGAGCATGCGAAGAAAGACAGCGGAGCTGGCCTTGGCACCGA 500
|||||
1921 uGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnGlySerV 1938
|||||
501 GGAGGAGCCACAGATGCTTAAGATGCTACAGATGTGCTTCAGGGCTCTG 550
|||||
1938 alG1yProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeu 1954
|||||
551 TGGGGCCCACTGAAACAGGGTCCCTGTGAGGTGCCAGGTGTTTTA 600
|||||
1955 AlaGluIleProGluAspProLysLeuPheArgHisAsnLysLeuArg 1971
|||||
601 GCAGAGATCCCGAAGACCCCAAGCTTCCCGCATCAACAACAAATTCG 650
|||||
1971 gLeuCysPheLysAspPheCysLysLysCysGluAspAlaLeuArgLys 1987
|||||
651 GCTGTGCTTCAAGACTTCTGCAAGAAATGTAGGATGCGTGGCGGAAC 700
|||||
1988 AsnLysAlaLeuIleG1yProAspGlnLysGluTyrHisArgGluLeuG1 2004
|||||
701 AATAAGGCCCTGATTGGCGCGAGCAAGAGAGTACCAACCTGAGACTGGA 750
|||||
2004 uArgAsnTyr.CysArgLeuArgGluAlaLeuGlnPro 2016
|||||
751 GCGCAACTCTTCCGCGCTGCGGAGGCTGTGCAGCCCC 788
|||||
seq_name: qb_est2:B1820302

```

```

REFERENCE 1 (bases 1 to 927)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11443 row: 0 column: 16
High quality sequence stop: 806.
FEATURES
source
1..927
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5178087"
/lab_host="DH10B"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gubler (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH-MGC Library."
BASE COUNT 214 a 257 c 292 g 164 t
ORIGIN
alignment_scores:
Quality: 1304.50 Length: 329
Ratio: 4.407 Gaps: 8
Percent Similarity: 89.970 Percent Identity: 87.234
alignment block:
US-09-736-968a-2 x B1820302 ..
Align seq 1/1 to: B1820302 from: 1 to: 927
1299 G1YrLysAlaPheGluArgG1LeaSerLeuThrPheLysSerIe 1315
|||||
2 GGGAAAAAGGCTTTTGAACCATCAACAGCCTCACATTCAAAAAATTTCT 51
|||||
1315 uAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrIleG1yAlaA 1332
|||||
52 GGATATGAAAGCGCGGCTAGAGAAAGCATTTCTGGTACCATGAGAGTC 101
|||||
1332 rGgGlnGluMetValArgArgSerArgGluArgSerProPheGlyAsnPro 1348
|||||
102 GACAAGAAATGGTTGGCGGAAGTCGTGAGAGAGCCCGTT. GGGAAATCCG 150
|||||
1349 GluAsnValArgTyrArgLysSerValThrHisTriPlyGlnThrSerAs 1365
|||||
151 GAGAAATGTCCGCTGGCGGAAGAGCGTCCACACTGGAGAACAACTCACA 200
|||||
1365 ArgValAspLysThrLysAspGlnMetGlnHisGluAlaLeuValGluG 1382
|||||
201 CCGGCTGACAAAGCAAGATGAATGGAACGAGAGCGCTTGCTGGAGG 250
|||||
1382 LysAsnLeuAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIle 1398
|||||
251 GGAACCTGGCAACCGAGGCAAGCCTAGTGTCTGACACACATCGAGATC 300
|||||
1399 IleValGlnThrValMetLeuSerGluAlaArgGluSerValLeuGluL 1415
|||||
301 ATCGTGCAGACGGTATGCTTTCAGAAACCCGGGAGAGCGCTTGGGGGC 350

```

```

1415 aValleuLysValValleuTyrSerleuGlySerAlaGlnSerAlaLeuP 1432
|||||
351 AGTGTGAAGGTTGTGTGTACAGCTGGGAGTGGCCAGAGTCCCTCT 400
1432 heLeuGlnHISGlyLeuAlaThrGlnArgAlaLeuValSerLysPhePro 1448
|||||
401 TCTTGACACATGGCTGGCCACCCAGAGGGCCCTGTGTGTCCAACTCCCG 450
1449 GluLeuLeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLeuAr 1465
|||||
451 GAGCTGCGTGTTCGAGGAGGACAGAGCTGTGTGCCGACCTGCTCCAG 500
1465 GluLeuLeuArgHisCysGlySerArgLLeSerThrLLeArgThrHisAlas 1482
|||||
501 GCTCTACAGACACTGTGGCAGCCGATCAGACATCCGACAGCAGCCCA 550
1482 eAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGlnLLeGlyHisAsn 1498
|||||
551 GGGCTCGCTGACTGCTCATGGCAGCAAGACTTCGAGATGGCCACAC 600
1499 PheAlaArgValLysMetGlnValThr.MetSerLeuSerLeuVal. 1514
|||||
601 TTGGCCCTGTGAAGATCAGAGTCAACATGTCTCTCTCTGCTGCTGG 650
1515 GlyThrThrGlnAsnPheSerGluGluHisLeuArgArgSerLeu. 1531
|||||
651 GGGACGACGACAGAACTTCAGTGAAGACACCTCGACCTCATCAAAA 700
1531 hrLeuLeuThrTyrAlaGluGluLysPheMetGlyLeuArgAspSerThrPh 1547
|||||
701 CCATCTTCACCTATGTCTGAGGAGCAGCAGGCGCGGACAGACCTT 750
1547 eAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetLLeuThrA 1564
|||||
751 CGCAGAGCAGAGTCCGAGACCTGATTTCACTGCATGATTCCTGACGG 800
1564 spThrValLysMetLys.GluHisGlnLysAspProGluMetLeuLLea 1580
|||||
801 ACACGCTGAAGATTCGAGGGGCCGAGAGGACCCCTGAGATT..... 843
1580 spLeuMetLysThrGlnLeuAlaArgLysGlnLysPheAspLeuArg 1596
|||||
844 .....GCTCCTCGGCTTCCA..... 858
1597 LeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnH 1613
|||||
859 .....TGTCCGACATGGCCGAGAGCAGCGGAGCTGGGGAACA 899
1613 sAlaGluAlaAlaGlnCysMetVal 1621
|||||
900 CGCCGAGGCGCCGACGTGCTGTGG 924
seq_name: gb_est2:BI830310
seq_documentation_block: 770 bp mRNA linear EST 04-OCT-2001
LOCUS BI830310 603073042F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5164762 5',
DEFINITION mRNA sequence.
ACCESSION BI830310
VERSION BI830310.1 GI:15941860
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 770)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: c9apbs-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.

```

```

FEATURES
    source
        cdna Library Preparation: Life Technologies, Inc.
        cdna Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
        DNA Sequencing by: Incyte Genomics, Inc.
        Clone distribution: MGC clone distribution information can be
        found through the I.M.A.G.E. Consortium/LNL at:
        http://image.lnl.gov
        Plate: LRAM11409 row: d column: 11
        High quality sequence stop: 760.
        Location/Qualifiers
            1..770
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone_IMAGE="5164762"
                /clone_lib="NIH_MGC_119"
                /tissue_type="medulla"
                /lab_host="DH10B"
                /note="organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: EcoRV (destroyed); RNA source normal medulla from
                anonymous male age 27. Library is oligo-dT primed and
                directionally cloned (EcoRV site is destroyed upon
                cloning). Average insert size 1.3 kb, insert size range
                0.9-3 kb. Library is normalized and enriched for
                full-length clones and was constructed by C. Gruber
                (Invitrogen). Research Genetics tracking code 013. Note:
                this is a NIH_MGC Library."
BASE COUNT 164 a 239 c 223 g 144 t
ORIGIN
alignment_scores:
    Quality: 1284.00 Length: 258
    Ratio: 5.035 Gaps: 2
    Percent Similarity: 98.837 Percent Identity: 98.450
alignment_block:
    US-09-736-968A-2 x BI830310 ..
Align seq 1/1 to: BI830310 from: 1 to: 770
77 GluPheProAlaAspAspLeuGluLeuLeuGlnProArgGluCysAr 93
|||||
1 GAATGCCACGTGATGACTTGAGAGCTGCTGTCAGGCCGCGGAATGCCG 50
93 gThrThrGluProGlyLLeProLysAspGluLysLeuAspAlaGlnValA 110
|||||
51 GACACGACGACCCCGGGATCCCAAGATGAGAAACGTGAGGCCACAGTGA 100
110 gAlaAlaValAlaGluMetLysLLeGluAspTrpValLLeValHisArgArg 126
|||||
101 GGGCCGCGGTGGAGATATATTTGAGGACTGGGTCAATTGCCACAAAG 150
127 TyGlnTyrLeuSerAlaAlaTyrSerProValThrThrAspThrGlnAr 143
|||||
151 TATCAGTACTGAGTGCAGCATACAGCCCGTCACACACACACAGCG 200
143 gGluArgGlnLysGlyLeuProArgGlnValPheGluGlnAspAlaSerG 160
|||||
201 GGACCGACAGAGAAGGGCTCCCGCCGAGGCTTTGGACAGAGATGCTTGG 250
160 LysAspGluArgSerGlyProGluLysPheSerAsnAspSerArgArgGlySer 176
|||||
251 GAGACGAGAGGTCCGGCCCTGAGGACTCGAATGACTCCCGGCTGGCTCG 300
177 GlySerProGluAspThrProArgSerSerGlyAlaSerSerLLePheAs 193
|||||
301 GGCTCCCGGAGAGACACCCCTCGAAGCAGTGGTCCCTGTGACATCTTGA 350
193 PheLeuArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArgA 210
|||||
351 CCTGAGGAACCTGCGACCTGACATTCATTCGCTGCTGAGACGGGG 400
210 lAlaProGluAspValAspArgArgAsnLThrLeuArgAlaGlnHis 226
|||||

```



```
401 CGGCCCCAAGAGATGTGGACCGCGCAATGAACCTTGACGCGCAGCAC 450
227 ATGProAlaLeuLeuThrLeuTyProAlaProAspGluAspGluAl 243
451 CGGCCCCCGGGCCCTGCTACCCCTCTACCCGCGCAGGATGAAC 500
243 aValGluArgCysSerArgProGluProProArgGluHisPheGlyLna 260
501 CGTGGAAACGTGTAGCCGCCCTAGAGACCAACCCCGAGACACTTGGACAA 550
260 rGllLeuValLysCysLeuSerLeuLysPheGluIleGluIleGluPro 276
551 GGATCTTGCTGAAGTGTCTGTGCTCAAGTTCGAGATTGAAATGAGACCC 600
277 lIepheGlylLeuAlaLeuTyArgPValArgGluLysLysLysIleSe 293
601 ATCTTTGGATCTTGCGCTGTATGATGTGGGAGAAAAAGATGATC 650
293 rGluAsnPheTyRPhAspLeuAsnSerAspSer.MetLysGlyLeuLeu 309
651 GGAAGAACTTCTACTTCGACCTGAACTCGACTCCATGAAGGGGCTCTT 700
310 ArGAlaHisGlyThrHisProAlaIleSerThrLeuAlaArgSerAla.I 326
701 CGGGCTCATGGGACCCACACCCGCGCATCTCCACCTGGCCGCTGCCCA 750
326 lPheSerValThrTyPro 332
751 TCTTCTCTGTGACTACCC 770
seq_name: gb_est2:BE531136
```

```
seq_documentation_block:
LOCUS BE531136 805 bp mRNA linear EST 09-AUG-2000
DEFINITION 601278315F1 NIH_MGC_39 Homo sapiens CDNA clone IMAGE:3610539 5',
mRNA sequence.
ACCESSION BE531136
VERSION BE531136.1 GI:9759781
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 805)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgrabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA library preparation: Ling Hong/Rubin Laboratory
DNA library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.jnl.gov
Plate: LCM266 row: m column: 04
High quality sequence stop: 770.
Location/Qualifiers
1..805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3610539"
/clone_11b="NIH_MGC_39"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: pancreas; Vector: pCMV7; Site: 1: XhoI;
Site: 2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
```

```
BASE COUNT 169 a 246 c 249 g 141 t
ORIGIN
alignment_scores:
Quality: 1256.00 Length: 266
Ratio: 4.906 Gaps: 2
Percent Similarity: 96.241 Percent Identity: 94.737
alignment_block:
US-09-736-968a-2 x BE531136 ..
Align seg 1/1 to: BE531136 from: 1 to: 805
1512 SetLeuValGlyThrThrGlnAsnPheserGluGluHisLeuAlaArgSe 1528
|||||
2 TCCCTGTGGGGGACGACGAGAACTTCAGTGAAGACACCTTCGACGTTCC 51
1528 rLeuLysThrIleLeuThrTyAlaGluGluAspMetGlyLeuArgAspS 1545
|||||
52 ACTCAAAACCATCCTACCTATGCTGAGAGAGACATGGGGCTGGCGGACA 101
1545 eThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIle 1561
|||||
102 GCACCTTCGACAGAGCAGTCCAGACCTGATGTTCACCTGCACATGATC 151
1562 lLeuThrAspThrValLysMetLysGluHisGlnGluAspProGluMetLe 1578
|||||
152 CTGACGCGACACGCTGACAGATGAGAGACACGAGAGACCTGAGATGCT 201
1578 uIleAspLeuMetTyArgIleAlaArgGlyTyGlnGlySerProAspL 1595
|||||
202 CATCGACCTCATGTATACAGAAATGGCCGGGGCTACACAGGGCTACCGGACC 251
1595 euArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGly 1611
|||||
252 TTCGGCTGACCTGTGTGCAGAACATGGCCGGGAAACACGCGAGCTGGGC 301
1612 AsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaLeuValAl 1628
|||||
302 AACACGCGCGAGCGCCGCCAGTGCATGTGTGCACGGCGCCCTCTGTGGC 351
1628 aGluTyRLeuAlaLeuLeuGluAspGlnArgHisLeuProValGlyCysV 1645
|||||
352 TGAGTACCTCGCCCTGCTGACAGACACCGCACCTGCCCGGGGCTCGC 401
1645 aISerPheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSer 1661
|||||
402 TTTCCTTCGAGAACATCTCATTCAAATGTGCTAGAGAGATCCGCCATCTCC 451
1662 AspAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPh 1678
|||||
452 GACGACATCTCTGCCCGCAGACAGAGAGGCTTCTCTCGGGAAACACTT 501
1678 eThrGluLeuGlyLeuValGlyLeuLeuGlnAlaAlaGlyTyRPhet 1695
|||||
502 CACTGAGCTGGGGGCTGTAGGGTCTGTGAACAGCAGCCGGCTACTTCA 551
1695 hrMetGlyClyLeuTyRgLuAlaValaLaengluValTyTyLysAsnLeuIle 1711
|||||
553 CCATGGGGGGGCTTACGAGGGGTGAATGAGGCTTCAAAACCTCATTC 601
1712 ProIleLeuGluAlaHisArgAspTyRysLysLeuAlaValaHisGly 1728
|||||
602 CCCATCTCGGAAGACCCACGCTGACTACAGAAAGCTGGCGGGGTGCACGG 651
1728 yLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerGlyTyRg 1745
|||||
652 CAAACTCGAGGAGGCTTCACCAAGATCATGCACAGAGTTCCGGCTGGA 701
1745 lUArgValPheGlyThrTyRPhetArgValGlyPheTyRgLyAlaHisPhe 1761
:: ::::
```

702 GC...GGCTGCGGACGATTCCGCTGGCTTACGGCGCCACTTC 748
1762 G1yAsPleuAsPgluGluPhuValTyrysluProserIleThr 1777
||||| |||||
749 GGTGACTG.GATGACGACGAG...TTGTGTCCAGAGGCTGCATCAGC 792
seq_name: gb_est2:B1854143

seq_documentation_block:
LOCUS B1854143 842 bp mRNA linear EST 10-OCT-2001
DEFINITION 603381452F1 NIH_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5389116 5',
mRNA sequence.
ACCESSION B1854143
VERSION B1854143.1 GI:15994890
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 842)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM1191 row: h column: 13
High quality sequence stop: 777.
Location/Qualifiers
1. 842
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="IMAGE:5389116"
/clone_id="NIH_CGAP_Mam4"
/tissue_type="tumor, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert 2.5 kb. Library constructed by Life
Technologies, catalog # 12018-016. Investigators providing
samples: Lothar Hennighausen/Priscilla Furth, NIH
Reference for transgenic model: Li et al., Cell Growth and
Differentiation 7, 3-11 (1996). Note: this is a NCI_CGAP
Library."
BASE COUNT 186 a 256 c 227 g 173 t
ORIGIN

alignment_scores:
Quality: 1237.00 Length: 284
Ratio: 4.633 Gaps: 7
Percent Similarity: 94.014 Percent Identity: 88.732

alignment_block:
US-09-736-968a-2 x B1854143 ..
Align seg 1/1 to: B1854143 from: 1 to: 842

242 G1uAlaValGluArgCysSerArgProGluProArg.GluHisPheG 258
||||| |||||
2 GAAGCTGTGAACCTGACGCCCTCGAGCCACCCCGATGAGCACTTGG 51
||||| |||||
258 TyGlnArgIleuValCysLeuSerIleuPheGluIleGluLe 274
||||| |||||
52 GACAGAGGATACATGTAAGTGGCTGTCACTCAAGTTTGAATTAATT 101
||||| |||||

275 G1uProIlePheG1yIleuAlaIeUtyrAspValArgGluTySly 291
||||| |||||
102 GAACCCATCTTCGGCACCTTGGCCCTGTATGATGTGGGAGAGAA 151
||||| |||||
291 sIleSerGluAsnPhetYrPheAsPleuAsnSerAspSerMetIly 308
||||| |||||
152 GATTTCAGAAATCTTCTTACTTTGACCTTAACCTCGGACTCCGTTAAGGGC 201
||||| |||||
308 euleuArgAlaHisGlyThrHisProAlaIleSerThrIleuAlaArgSer 324
||||| |||||
202 TGCCTCCGGGCCCATGGACACACACACCATCTCAGCTTGGCAGGCTCT 251
||||| |||||
325 AlaIlePheSerValThrTyProSerProAspIlePheIleuValIely 341
||||| |||||
252 GCCATCTTCTCCGTGACCTACCCATCTCTGACATCTTCTCGTTGCCAA 301
||||| |||||
341 sIeuGluTyValIeuGlnGlnGlyAspIleSerGluCysCysGluPro 358
||||| |||||
302 GCTGGAAAGGTGCTGCACGACGAGGGGACATCAGTGAAGTCTGTAACCT 351
||||| |||||
358 yrMetValIeuTyGluValAspThrAlaIleAsnTyGluTyLeuGlu 374
||||| |||||
352 ACATGCTGATGAGAGGACAGACACAGCAAGAAAGAAAGAGCTGGAG 401
||||| |||||
375 LysIeuArgLeuAlaIaGluGlnPheCysThrArgIleuTyArgTyArg 391
||||| |||||
402 AAGCTGCGCTGCTGCGACGACATCTGACACGAGCTGGAGCGCTACCG 451
||||| |||||
391 gMetProPheAlaThrPhrAlaValHisLeuAlaAsnIleValSerSera 408
||||| |||||
452 CATGCCCTTCCCTGGAGCGGTGTGCACTTGGCCCAATGTGTGACGG. 500
||||| |||||
408 laGlyGlnLeuAspArgAspSerAsp.SerGluGlyGluArgArgProAl 424
||||| |||||
501CCCCAGACCGGAGATTCGCACTCTGAGGCGCAACGACACCCAC 545
||||| |||||
424 aTrpThrAspArgArgArgArgGlyProGlnAspArgAlaSerSerGlyA 441
||||| |||||
546 CTGGGCCCAACCGCCGCTGCGGGGCCCCAGAGATGAGGCTATGTGGG 595
||||| |||||
441 sPAspAlaCysSerPheSerGlyPheArgProAlaThrIleuThr.ValTh 457
||||| |||||
596 ATGATGCTGTGACGCTTCTTACGCTTCCGCGCGGCGCACGTTAATGTTCAC 645
||||| |||||
457 rAsnPhenylsclngluAlaGluArgLeuSerAspGluAspLeuPheL 474
||||| |||||
646 CAACCTTTTCAACAGCAGGAGCAGAACGCTCAGATGAAGAACCTTTTCA 695
||||| |||||
474 yspPheLeu.AlaAspMetArgArgProSerSerIleuLeuArgArgLeuAr 490
||||| |||||
696 AGTTTCTGGCCGACATGCGGCGACCATCATCTCCTGTGAGGCGAAATGCG 745
||||| |||||
490 gProValThrAlaGln.LeuTyIleAspIleSerProAlaProGluAsn 506
||||| |||||
746 CTCTGTGACAGCTCAGCCTT.....GACATCTCCCGCGCGCCGAGAAC 789
||||| |||||
507 ProHisPheCysLeuSerProGluLeuLeuHisIleTyPro 520
||||| |||||
790 CTGCACCTTCTGCTGCTCCCGATTAACCTTCATGTTCAGCCT 831
||||| |||||

seq_name: gb_est2:B1078974

seq_documentation_block:
LOCUS B1078974 777 bp mRNA linear EST 20-JUN-2001
DEFINITION 602873230F1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:5005183 5',
mRNA sequence.
ACCESSION B1078974
VERSION B1078974.1 GI:14497304
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE      1 (bases 1 to 777)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                  Email: cgapsb@remail.nih.gov
                  Tissue Procurement: Gilbert Smith, Ph.D.
                  cDNA Library Preparation: Life Technologies, Inc.
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLM11044 row: k column: 08
                  High quality sequence stop: 739.

FEATURES
  source
    1..777
    /organism="Mus musculus"
    /strain="FVB/N-3"
    /db_xref="taxon:10090"
    /clone_image="5005183"
    /clone_id="NCL_CGAP_Mam2"
    /tissue_type="tumor, biopsy sample"
    /dev_stage="5 months"
    /lab_host="DH10B"
    /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
    Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
    Library constructed by Life Technologies. Investigator
    providing samples: Gilbert Smith, NIH"

BASE COUNT      182 a      213 c      222 g      160 t
ORIGIN
alignment_scores:
  Quality: 1213.00      Length: 260
  Ratio: 4.720          Gaps: 2
  Percent Similarity: 98.846 Percent Identity: 96.154

alignment_block:
US-09-736-968a-2 x BI078974 ..
Align seg 1/1 to: BI078974 from: 1 to: 777

1483 AAlasertleuTyrleuLeuMetArglnAsnphglnleuGlnHisasnpH 1499
|||||
2 GCGTCCCTCTACTGCTTATGCGCAGAACTTCGAGATTGGCCATTAACCTT 51
1499 eAlaArgValysmetGlnValThrMetSerLeuSerSerLeuValGlyT 1516
|||||
52 TCCCGGTGGAGAGATGCTGTGACCATGCTCTGTCGCTTGTGGGGA 101
1516 hTThGlnAsnPheserGlnHisLeuArGArGSerLeuLysThrIle 1532
|||||
102 CAACTCAGAAACTTTAGTGAAGGACATTTGAGAAAGTCCCTCAAGACCATC 151
1532 leuThrrTyrAlaGlnGluAspMetGlyLeuArGArGSpSerThrPheAlaG 1549
|||||
1533 leuThrrTyrAlaGlnGluAspMetGlyLeuArGArGSpSerThrPheAlaG 1549
|||||
152 CTGACCTTACGAGAGAGACATAGGCTGTGAGGACAGACACCTTCGCTGA 201
1549 uGlnValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrV 1566
|||||
202 GCAGGTCCAGAGACCTCATGATCTCAACCTGCACATGATCTGCAGACACGG 251
1566 aLysMetLysGlnHisGlnLusApproGluMetLeuLaspLeuMet 1582
|||||
252 TGAAGATAGAGGAACACAGAGGACCTGAGATGCTATGAGACTCATG 301
1582 TyrArgGlleAlaArgGlyTyrGlnGlySerProAspLeuArGLeuThrTr 1599
|||||
302 TACACAATTTGCTCGGGGATACCAAGGCTCTCAGACCTGCGCGTCGATG 351
1599 PLeuGlnAsnMetAlaGlyLysHisAlaGlnLeuGlnLysHisAlaGlnU 1616
|||||
352 GCTCAGAACATGGCTGGCAAAACATGCAGAGCTGGGCATATCATGCAGAG 401
1616 laAlaGlnCysMetValHisAlaAlaAlaLeuValAlaGluTyrLeuAla 1632
|||||
402 CTGCCAGATGATGATGTCACAGCGGCTGCGGTGGCGGAATACCTTGGC 451
1633 leuLeuGlnAspGlnArGHisLeuProValGlyCysValSerPheGlnAs 1649
|||||
452 CTCCTCAGAGCACTGCCACCTGCTGTGGCTGTGTCTTCCCGAA 501
1649 nIleSerSerAsnValLeuGlnGlnSerAlaIleSerAspIleLeuS 1666
|||||
502 TGTCTCATTCACAGCTGCTGAGAGAGCTGCCATCTCTGATGACATCTGT 551
1666 erProAspGlnGluGlyPheCysSerGlyLysHisPhehrGluLeuGly 1682
552 CGCCAGATGAGAGAGGCTTCTGCTCTGGAAGAACCTTCACAGAACTGGGT 601
1683 leuValGlyLeuLeuGlnGlnAlaAlaGlyTyrPhehrMetGlyGlyLe 1699
|||||
602 CTGTGTGGGCTGCTGGAGAGGCGCGGCTACTTCAACATGGGTGCTT 651
1699 uTyrGlnAlaValAsnGlnValTyrLysAsnLeuIleProIleLeuGlnU 1716
|||||
652 GTATGAAGCGGCTGACAGAACTTTCMAAAACCTTATCCCATCTTGAAG 700
1716 laHisArgAspTyrLysLysLeuAlaValHisGlyLysLeuGlnGlu 1732
|||||
701 C.CACAGAGACTTCAAGAACCTGTGT.GCGGTGCACGGGAACATCGAGAG 749
1733 .AlaPheThrLys.IleMetHisGln 1740
|||||
750 GGCTTCACCCAGAGATATATCACCAG 775

seq_name: gb_est2:BI181383
seq_documentation_block:
LOCUS      BI181383      769 bp      mRNA      linear      EST 10-JUL-2001
DEFINITION UNL-P-FN-ah-e-08-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
ACCESSION  UNL-P-FN-ah-e-08-0-UNL.3', mRNA sequence.
VERSION    BI181383
KEYWORDS   BI181383.1 GI:14655792
SOURCE     EST.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 769)
AUTHORS    Caetano, A.R., Johnson, R.K. and Pomp, D.
TITLE      Generation and sequence characterization of a normalized cDNA
            library from swine ovarian follicles
JOURNAL    Unpublished (2001)
COMMENT    Contact: Pomp, D
            Department of Animal Science
            University of Nebraska, Lincoln
            Lincoln, NE 68583-0908, USA
            Tel: 402 472 6416
            Fax: 402 472 6362
            Email: dpompe@unl.edu
            Oligo-dt track not found, Not I site shown in beginning of sequence
            is likely internal to the message.
            Seq primer: M13 -29
            POLYA-No.

FEATURES
  source
    1..769
    /organism="Sus scrofa"
    /strain="University of Nebraska, Lincoln Swine Selection
    lines"
    /db_xref="taxon:9823"
    /clone_image="UNL-P-FN-ah-e-08-0-UNL"
    /clone_id="UNL-P-FN"
    /dev_stage="ADULT"
    /lab_host="DH10B (Life Technologies)"

```

/note=Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; The UNL-P-FN
library is a normalized library representing porcine
ovarian follicles, ranging between 2.0 to 10.0 mm in
diameter, collected during 7 days of the follicular phase
of the pig estrous cycle. This library was derived from
the library UNL-P-F2. The tag is a string of 5-6
nucleotides present between the Not I site and the
oligo-dT track. The library was constructed as described
by Bonaldi, Lennon and Soares, Genome Research 6: 791-806
, 1996.
TAG_SEQ=None found"

BASE COUNT 162 a 246 c 223 g 137 t 1 others
ORIGIN

alignment_scores:
Quality: 1211.00 Length: 252
Ratio: 4.903 Gaps: 0
Percent Similarity: 98.016 Percent Identity: 92.063

alignment_block:
US-09-736-968a-2 x B1181383 ..

Align seg 1/1 to: B1181383 from: 1 to: 769

```
112 A1aValGluMetTyrTlEgluAspTrrValIleValHisArgTyrG1 128
|||||
11 GCGGTGGAGATGTACACCGAGCTGTGATCACCACAGAGAGTACCA 60
|||||
128 nTyrLeuSerAlaAlaTyrSerProValThrThrAspThrGlnArgGua 145
|||||
61 GCACCTGAGTGCAGTGTACAGCCCACTACTCAGAGACACGACGAGAC 110
|||||
145 rGgInLysGlyLeuProArgGlnValPheGluGlnAspAlaSerGlyasp 161
|||||
111 GGGAGAAAGGCGCTCACCGGAGGTCTTGACAGAGATGCTTGGGAGT 160
|||||
162 GluArgSerGlyProGluAspSerAsnAspSerArgArgGlySerGlySe 178
|||||
161 GAGAGGTCTGGCCCAAGAGACTCGGTGACCCCGCAGCTCCTCAGGCTC 210
|||||
178 rProGluAspThrProArgSerSerGlyAlaSerSerIlePheAspLeu 195
|||||
211 CCCAGATGACACCCACGAGACATGGCGCTCTGACATCTTTGACCTGA 260
|||||
195 rGAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArgAla 211
|||||
261 GGAACCTGGCGGCGGACGCTCTGCTGCCCTCACTGCTGAGGCGTGGCC 310
|||||
212 ProGluAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgPr 228
|||||
311 CCAGAGAGAGTGGACCGCGCAATGAGGCTCTGACGACGACACCGCGCC 360
|||||
228 oProAlaLeuLeuThrLeuTyrProAlaProAspGluAspGluAlaValG 245
|||||
361 CCGCGCCCTGCTACCTCTACCGACCTGATGAGGAGGAGGCTGTGG 410
|||||
245 LArgGlySerArgProGluProProArgGluHisPheGlyGlnArgLe 261
|||||
411 AACGCTGACGCGCGCGGAGCCACCCGAGAACTCTGCGACACAGGATC 460
|||||
262 LeuValLysCysLeuSerLeuLysPheGluIleGluIleGluProIlePh 278
|||||
461 CTGGCTCAAGTGTCTGTCTTAAGTTCGAGATCGAATCGAGCCCATCTT 510
|||||
278 eGlyIleLeuAlaLeuTyrAspValArgLysLysLysIleSerGlu 295
|||||
511 TGGCATCTTAGCCCTGTATGATGTTCTGGGAAAAAAGAGATCTCAGAGA 560
|||||
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
|||||
561 ACTTCTACTTCGACCTGAACTCAGATCATGAAGGGCTGCTGGCGGCC 610
```

seq_name: gb_est2:BG678867
seq_documentation_block: 857 bp mRNA linear EST 01-MAY-2001
LOCUS BG678867
DEFINITION 602624767F1 NCI_CGAP_Skn4 Homo sapiens cDNA clone IMAGE:4749689 5',
mRNA sequence.
ACCESSION BG678867
VERSION BG678867.1 GI:13910264
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 857)
TITLE NIH-MGC http://mgc.nci.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: James Cleaver, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL0603 row: e column: 18
High quality sequence start: 4
High quality sequence stop: 818.
Location/Qualifiers
1. 857
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4749689"
/clone_11b="NCI_CGAP_Skn4"
/tissue_type="squamous cell carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: skin; Vector: pCW-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5kb. Library constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 209 a 254 c 247 g 147 t
ORIGIN

alignment_scores:
Quality: 1193.00 Length: 270
Ratio: 4.715 Gaps: 7
Percent Similarity: 93.704 Percent Identity: 91.111

alignment_block:
US-09-736-968a-2 x BG678867 ..

Align seg 1/1 to: BG678867 from: 1 to: 857

1784 HisArgLeuGluGluPheTyrThrGluArgPheGlyAspAspValValG1 1800

```

|||||
4 CACGGCGTGGAGAGTTCTACACGAGAGATTGGCGACGCGCTTGA 53
1800 uil1el1e1ysaspseryrprovalasplyserslysleuaspsenl 1817
54 GATTATCAAAAGACTCTAACCTGTGGCAAGTCCAACTGACTGACACAAA 103
1817 ysala1y1le1g1n1e1th1y1val1u1p1ro1y1r1heas1p1h1y1g1u 1833
104 AGGCTTACATCCAGATCACCATATGGAACCGTACTTGTACTTACGAG 153
1834 leulysaspargval1th1y1r1heas1p1r1g1s1y1leu1a1r1g1th1r1p 1850
154 CTCAAGACCGGGGTGACCTACTTGGACCGCAACTATGGCTTGCACATT 203
1850 eleuPhecys1th1r1p1he1r1p1ro1as1p1s1y1a1r1g1u1h1s1g1y1u1e1u1p 1867
204 CTTCTTTCGACGCGCTTCACGCGGATGGCGCGACGCGGAGCTGC 253
1867 rogl1n1h1s1y1a1r1g1s1y1th1r1leu1e1u1s1er1th1r1as1p1h1s1a1l1a1p1he1r1p 1883
254 CCGAGCAACACAAAGCTTAAGCGCTGCTCAGCAACGACGCGCTTCCC 303
1884 Ty1r1le1y1s1th1r1a1g1l1e1a1r1g1a1l1y1s1h1s1a1r1g1u1h1u1th1r1a1l1e1u1th 1900
304 TACATCAAGACTCGCATCCGTGTGTCACCGGAGAGACGCTGTGAC 353
1900 r1p1ro1a1l1y1a1l1a1r1l1e1g1u1a1s1p1m1e1t1s1y1l1y1s1th1r1a1r1g1u1e1u1a 1917
354 GCCAGTGAGAGTGGCCATCAAGACATGCAAGAAAGAACACGGGAGCTGG 403
1917 la1p1he1a1r1h1g1u1g1n1a1s1p1r1o1a1s1p1a1l1a1y1s1m1e1t1e1u1g1m1e1t1a1l 1933
404 CCTTGCACCGACGACGACGCCACAGACTCTAAGATGCTACAGATGTC 453
1934 leu1g1n1g1y1s1e1r1a1l1y1p1r1o1th1r1a1s1n1g1y1p1r1o1e1u1l1u1a1l1 1950
454 CTTAGGCGCTGTAGGGCCACCGTGAACCA.GGTCCCTGGAGGTGGC 502
1950 a1g1n1a1l1p1he1u1a1l1u1l1e1p1r1o1g1u1a1s1p1r1o1l1y1s1e1u1p1h1e1a1r1g1h1s1h 1967
503 CCAGCTTTTGTAGCAGAGATCCGGGAAGACCCCAAGCTTCCGGATC 552
1967 l1s1a1n1y1s1e1u1a1r1g1e1u1c1y1s1p1h1e1y1a1s1p1h1e1c1y1s1y1s1c1y1s1g1u1a1s 1983
553 ACAACAAATTGGCGCTCTCTTCAAGACTTCTGCAAGAAATGTGAGGA 602
1983 p1a1l1e1u1a1r1g1y1s1a1n1y1s1a1l1e1u1l1e1g1y1p1r1o1a1s1p1g1n1y1s1g1u1t1y1r1h 2000
603 TGGCTTCGGGAAAAATTAAGCCCTGATGGGGCCGGACCAAGAGATACC 652
2000 l1s1a1r1g1u1e1u1g1u1a1r1g1a1s1n1y1r1c1y1a1r1g1e1u1a1r1g1u1a1l1e1u1g1n 2015
653 ACCGGGAGCTTGAAGCGCAACTATGCCCTGGGGAGGCTGTGAG 702
2016 p1r1o1e1u1e1u1h1r1g1n1a1r1g1e1u1p1r1o1g1n1e1u1m1e1t1a1l1p1r1o1h1r1p1r1o1 2030
703 CCTGCTTACCCCAAGCGGCTGCCCCACTAATGGCAATGGCCACCCCA 752
2031 .p1r1o1g1y1l1e1u1a1r1g1a1s1n1s1e1r1e1u1a1s1n1a1r1g1a1s1e1r1p1h1e1a1r1g1y1s1a1a1s1p 2046
753 GCCAGGCTTCAAGACTCTTGGACACAGCAAGTTTCCCGAAGGGAG 802
seq_name: gb_est2:BI159195
seq_documentation_block:
LOCUS BI159195 778 bp mRNA linear EST 05-JUN-2001
DEFINITION 602922704P1 NIH_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5063064 5',
ACCESSION BI159195
VERSION BI159195.1 GI:14619196
KEYWORDS EST.
SOURCE house mouse.

```

```

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 778)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1170 row: g column: 01
High quality sequence stop: 750.
location/Qualifiers
1..778
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5063064"
/clone_lib="NIH_CGAP_Mam3"
/tissue_type="tumour, gross tissue"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dr.
Average insert 2 kb. Library constructed by Life
Technologies, catalog #12017-018. Investigators providing
samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference
for transgenic model: Xu et al., Nature Genetics 22, 37-43
(1999). Note: this is a NIH_CGAP Library."
BASE COUNT 173 a 223 c 226 g 156 t
ORIGIN
alignment_scores:
Quality: 1178.00 Length: 262
Ratio: 4.656 Gaps: 4
Percent Similarity: 96.565 Percent Identity: 93.511
Alignment_block:
US-09-736-968a-2 x BI159195 ..
Align seg 1/1 to: BI159195 from: 1 to: 778
1373 G1UMeTG1uH1aSG1uAla1eUvAlG1uG1yA1sN1eUa1aThrG1uAla1e 1389
|||||
1 GAAATGAGACATGAAAGCTGTGTGAGCGGAACCTGGCGACAGAGCCAG 50
1389 r1eUa1vAl1eUa1s1p1h1r1eU1g1u1l1e1a1l1e1a1l1y1a1l1e1u1a1s1e1u 1405
|||||
51 CCTGTGTGTGTGAGACAGCGCTGTGAGACATCGTGCACAGAGTGTGTG 100
1406 s1e1r1g1u1a1a1h1a1g1u1s1e1r1vAl1eU1g1y1a1vAl1eU1y1s1y1a1l1eU1t 1422
|||||
101 TCCGAGGCCCGGTGAAGAGACTCTGTAGTCCGTGCGAATGTCTTCCCT 150
1422 y1s1e1r1e1u1g1s1e1r1a1g1n1s1e1r1a1l1a1p1h1e1u1g1n1h1s1g1y1l1eU1 1438
|||||
151 ACAGCTTGGAGAGCCCGAGAGTGCCTGTCTTCTGCAACATGGCCTGGC 200
1438 a1Th1r1n1a1r1g1a1l1eUa1s1e1r1y1s1p1h1e1r1o1g1u1e1u1p1h1e1u1g1u1a 1455
|||||
201 CACCGAGGGGGCCCTGCTCCAAAGTTCCGGAGCTGTTTTCAGAGAG 250
1455 s1p1Th1g1u1e1u1c1y1a1a1s1p1e1u1c1y1s1e1u1a1r1g1e1u1a1r1h1s1c1y1s1g1y 1471
|||||
251 ACACGAGCTTGTGTCCGACCTGTGCTGAGAGACTTCTGCGACACTGTGGC 300
1472 s1e1r1g1l1s1e1r1h1l1e1a1r1g1h1s1a1s1e1r1a1s1e1r1e1u1y1r1l1eU1 1488
|||||

```

```

301 AGCCGATCAGACACATCCGATGCACGCGCCTCCCTACCTACCT 350
1468 UMEkArGlnAsnPheGluileGlyHisAsnPheAlaArgValIysMetG 1505
351 TRATGGCCAGAACTTCGAGATTGGCCATACCTTCCTCCGTGTGAAGTGC 400
1505 InValTrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPheSer 1521
401 TGTGTACCATGTCTCTGTGCTCCCTTGTGTGGGACAACCTCAGACTTAA 450
1522 GUGUuHtStLeuAArgSerLeuLysThrIleLeuThrTrTyRAlaGluG 1538
451 GAAGAGCATTTGACAAAGTCCCTCAAGACCAATCCGACTCGAGAGAGA 500
501 GGACATAGGGGTGAGGAGACACCTTCCTGAGCAGGTCCGAGACCTCA 550
1555 eLPheAsnLeuHISMetIleLeuThrAsPThrValLysMetLysGluHis 1571
551 TGTTCACCTGTCACATGATCTTGACAGACAGCGTGAAGAATGAAGAACAC 600
1572 GlnGluAsPProGluMetLeuIleAsPLeuMetTyRArgIleAlaArgG 1588
601 CAGGAGGACCTGTGATGCTCATGAGACCTCATGATGACAGATCGCTGG 650
1588 yTyRGlnGlySerProAsPLeuAArgLeuThrTrPLeuGlnAsnMetAlaG 1605
651 ATACCAATGCTCTCCAGACTGCGCTGACCTGCGCTGCGAGAACATGGCTG 700
1605 IyLysHISAlaGluLeuGlyAsnHISAla.GluAlaAlaGlnCysMetVa 1621
701 GCAAAACATGCGAGACTGGGCATCATGCAAGAGGCTGCCCATGTCATGT 750
1621 IHISAlaAlaAlaValAlaGluLysTr 1630
751 CCACGCGCTGCGCTGCGTGGCGAATAC 778

seq_name: gb_est1:AW742035

seq_documentation_block:
LOCUS AW742035 692 bp mRNA linear EST 07-SEP-2000
DEFINITION up52c04.y1 Soares mouse NMIE Mus musculus cDNA clone IMAGE:2779686
5' similar to WP.F46H5.4 CE04590 RAT TRG GENE PRODUCT ;, mRNA
sequence.
ACCESSION AM742035
VERSION AM742035.1 GI:7653891
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Murinae; Mus.
1 (bases 1 to 692)
TITLE NCBI-GAP http://www.ncbi.nlm.nih.gov/ncigap.
NATIONAL CANCER INSTITUTE, Cancer Genome Anatomy Project (CGAP),
TUMOR Gene Index
Unpublished (1997)
Other ESTs: up52c04.x1
Contact: Robert Strausberg, Ph.D.
Email: cgrahbs-remail.nih.gov
This clone is available royalty-free through LNL : contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: 40RP from GlpDco
High quality sequence stop: 473.
Location/Qualifiers
1..692
/organism="Mus musculus"
/strain="C3H x 101 (F1 stock)"
/db_xref="taxon:10090"
/clone="IMAGE:2779686"
/clone_lib="Soares_mouse_NMIE"
/sex="male"
/dev_stage="newborn"

```

/lab.host="DHI08"
 Note: *Organ: inner ear, 170 pooled. Vector: pT73D-Pac;
 Site_1: Not 1; Site_2: Eco RI; 1st strand cDNA was primed
 with a Not I - oligo(dT) primer [5'
 TGTAAACCAATCTGAAAGTGGAGGCGGCGACACTTTTCTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified pT73 vector. Library
 is normalized, and was constructed and donated by Bento
 Soares and M. Fatima Bonaldo (University of Iowa) and R.
 Hardisty, A. Varela-Carver, P. Mburu and S.D.M. Brown (MRC
 UK Mouse Genome Centre and Mammalian Genetics Unit,
 Harwell, UK).

BASE COUNT	183 a	186 c	188 g	133 t	2 others
ORIGIN					

Alignment--scores:
 Quality: 1151.00 Length: 230
 Ratio: 5.161 Gaps: 0
 Percent Similarity: 96.957 Percent Identity: 94.348

alignment_block:
 US-09-736-968A-2 x AM742035 ..

Align seg 1/1 to: AM742035 from: 1 to: 692

1778	LysleuAlaGluIleSerHisArgLeuGluGluPheTyrrHgluAtrph	1794
3	AAGCTTCGAGGATCTCACACGGCTGGAGGAGTTCTATACGGAAAGTT	52
1794	eGlYAspAspValValGluIleIleLysAspSerTyrrProValAspLys	1811
53	CGGGAGATGTGTGTACAGATCATCAAGATTCTAACCGAGTGACAGT	102
1811	eLysLeuAspSerGlnLysAlaTyrlIleGlnIleThrTyrrValGluPro	1827
103	CCAGGCTGGACCCACAGAAAGCGTACATACATAACTCATGTGGAGCA	152
1828	TyrPheAspThrTyrrGluLeuLysAspArgValIthrTyrrPheAspArg	1844
153	CATTTCGCACCTTATGAGCTCAAGGATGCGGTACTTACTTCGATCGGAA	202
1844	nTyrGlyLeuArgThrPheLeuPheCysThrProPheThrProAspGly	1861
203	CTATGGGTGGGGCTCTCTCTTCTGCAACCCCTTCCACACAGATGGAC	252
1861	rgAlaHisGlyGluLeuProGluGlnHisLysArgLysThrLeuLeuSer	1877
253	GTGGCGACGCGAGATTGGCCGAACACACAAACGCAAGCCGTGCTGAGC	302
1878	ThrAspHisAlaPheProTyrrIleLysThrArgIleArgValCysHisArg	1894
303	ACGGAGCATGCTCTTCCCTACATCAAGACACGCGATCCGAGTGTGCCACG	352
1894	gGluGluThrValLeuThrProValGluValAlaIleGluLysPheGlnL	1911
353	TGAGGAGACAGTGGCTGACACCACTGGAGGTGGCCATTGAGGACATGCA	402
1911	ysLysThrArgGluLeuAlaPheAlaThrGlnGluAspProProAspAla	1927
403	AGAAAGACCCGGGAGCTGGCTTTGCCACCGACGAGAGCCCTCCAGATGCC	452
1928	LysMetLeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGly	1944
453	AAGATGCTGCGAGATGTTCTCCAGGGTCTCTGGGAGCCACCATGTGAACA	502
1944	ngLProLeuGluValAlaGlnValaPheLeuAlaGluIleProGluAsp	1961
503	NGGTCCCTGGAAATGGCCANGTGTTTTTGTCAAGATCCAGAAAGATC	552
1961	gLyLysLeuPheArgHisHisAsnLysLeuArgLeuCysPheLysAspPhe	1977

```
553  CCAAGCTCTTCGACATCACAACTCCGGCTCTGCTTCAAGATTTC 602
1978  CysIysIysCysGluAspAlaLeuArgIysAsnIysAlaLeuIleGIYPr 1994
      |||||
603  TGCATAAAGTGCAGAGATGCACTGAGAAAGACAAAGCCCTGATTGGGCC 652
1994  oAspGlnIysGluTyrHisArgGluLeuGluArgAsnTyr 2007
      |||
653  CAACCCAGATGATACCCACCGGAGAGCTTGAGCGTCACTAT 692
```


CC useful for detecting the CLASP-7 polypeptide. (ii) is useful for
CC inhibiting (i) by recombinant methods. (i) or its fragment are useful for
CC producing an immune response in a cell such as T cell or B cell. A
CC pharmaceutical composition (C), comprising (i) or (ii), can be useful
CC for treating CLASP-7-mediated disease such as an autoimmune disease
CC caused or exacerbated by increased activity of T helper cells. Autoimmune
CC diseases which can be treated using (C) include multiple sclerosis,
CC juvenile diabetes and rheumatoid arthritis. (i) is useful for treating
CC toxemia or pregnancy induced hypertension, pruritic urticarial papules
CC and Rh incompatibility. (i) is also useful as a diagnostic reagent for
CC immune and other disorders, since diseases characterised by
CC overproduction or depletion of lymphocytes in blood or other organs may
CC be detected by monitoring the level of (i) or its mRNA. CLASP-7 has been
CC mapped to the chromosomal location 19q13.2. The present sequence
CC encodes the human CLASP-7 protein as given in the present invention.
XX
SQ

Sequence 6372 BP: 1273 A; 2050 C; 1849 G; 1200 T; 0 other;

alignment_scores:

Quality: 10569.00 Length: 2047
Ratio: 5.163 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-736-968a-2 x AAH43851 ..

Align seg 1/1 to: AAH43851 from: 1 to: 6372

1 MetAlaLaSerGIuArGAlaPheAlaHisLysIleAsnArgThrVa 17
|||||
13 ATGCTGCTCCGACGCGCGCTTCGCGCACAAAGATCAACAGACGCGT 62
17 LAAlaIaGIuValArgLysGIuValSerArgGIuArgSerGIySerProH 34
|||||
63 GCGCGAGAGCTGCGAGAGAGGTGTCCGCGAAACGACGTGCTCCGCC 112
34 IsSerSerArgArgCysSerSerSerLeuGIyValProLeuThrGIuVal 50
|||||
113 ACTCCAGACAGCGCTGACAGAGCTCCCTGGGGGTCCCACTGACTGAAGT 162
51 ValGIuProLeuAspPheGIuAspValLeuLeuSerArgProProAspAl 67
|||||
163 GTCGAGGCCCTTGACATTGAGATGTACTTCGACCGCGCCACCAAGATGC 212
67 aGIuProGIyProLeuArgAspLeuValGIuPheProAlaAspAspLeuG 84
|||||
213 TGAAGCCGGGGCCCTCAGGAGCTGTAGAAATCCCAAGCTGATGACTTGG 262
84 LuLeuLeuLeuGIuProArgGIuCysArgThrThrGIuProGIyLeuPro 100
|||||
263 AGCTGCTGCTGCAGCCCGGGAAATGCCGACACAGGAGCCGGGATCCCC 312
101 LysAspGIuLysLeuAspAlaGIuValArgAlaAlaValGIuMetTyrI1 117
|||||
313 AAGGATGAAAAAAGTGGATGCCAGGTGAGGGCCGCGTGAGATGTATAT 362
117 eGIuAspTrValI1eValHisArgArgTyrGIuTyrLeuSerAlaAlaT 134
|||||
363 TGAAGGATGGGTCTATTGTCACAGAAAGTATCACTGAGTGCAGCAT 412
134 YrSerProValI1ThrAspThrGIuArgGIuArgGIuLysGIyLeuPro 150
|||||
413 ACAGCCCGCTGACACAGACACACAGCGGAGGAGACAGAAAGGCTCCCC 462
151 ArgGIuValPheGIuGIuAspAlaSerGIyAspGIuArgSerGIyProGI 167
|||||
463 CGCCAGGTCTTTGAGCAGATGCTCTGAGAGACAGAGGTCCGCCCTGA 512
167 uAspSerAspAspSerArgArgGIySerGIySerProGIuAspThrProA 184
|||||
513 GGAATCGAAATGACTCCGCGGCTGGGCTCCCGGAAAGACACCCCTC 562

184 YrSerSerGIyAlaSerSerI1ePheAspLeuArgAspLeuAlaAlaAsp 200
|||||
563 GAACAGAGGTGGCTTACATCTTCACCTGAGAGAACTGGCAGCTGAC 612
201 SerLeuLeuProSerLeuLeuGIuArgAlaAlaProGIuAspValAsp 217
|||||
613 TCATTGCTGGCCCTTCGCTAGACGGGGGGCCGCCAAGATGTGGACG 662
217 GArAsnGIuThrLeuArgArgGIuHisArgProProAlaLeuLeuThrL 234
|||||
663 GCGCAATGAACCTTTCGACGCGACGACCGGCCCGGCTGCTACACC 712
234 eUTyrProAlaProAspGIuAspGIuAlaValGIuArgCysSerArgPro 250
|||||
713 TCTACCCGGGACCTGACGAGATGAACCGTGAACCTGTAGCCGCCCA 762
251 GIuProProArgGIuHisPheGIuArgI1eLeuValLysCysLeuSe 267
|||||
763 GAGCCACCCCGGACACTTGGACAAAGATCTTGGTCAAGTGTCTGTC 812
267 rLeuLysPheGIuI1eGIuI1eGIuProI1ePheGIyI1eLeuAlaLeuT 284
|||||
813 GCTCAAGTTGAGATTGAAATTGAGCCCATCTTGGGATCTGGCTCTGT 862
284 YrAspValArgGIuLysLysLysI1eSerGIuAsnPheTyrPheAspLeu 300
|||||
863 ATGATGTGCGGAGAAAAAGAAATCTCGAGAACTTCACTTCGACCTG 912
301 AsnSerAspSerMetLysGIyLeuLeuArgAlaHisGIyThrHisProAl 317
|||||
913 AACTCGGACTCCATGAAGGGGCTCTTCGGGCTATGACACCCACCTGCG 962
317 aI1eSerThrLeuAlaArgSerAlaI1ePheSerValThrTyrProSerP 334
|||||
963 CATCTCCACCTTGCCCGCTGCTGACATCTTCTGTACCTACCCCTCAC 1012
334 rOAspI1ePheLeuValI1eLysLeuGIuLysValLeuGIuGIuLysAsp 350
|||||
1013 CTGACATCTTCCTGCTCATCAAGTTGGAGAGGTCTTCAGCAAGGGGAG 1062
351 I1eSerGIuCysCysGIuProTyrMetValLeuLysGIuValAspThrAl 367
|||||
1063 ATCACTGAGTCTGAGACCTTACATGGTGTGAAAGAGTGGACACACACC 1112
367 aLysAsnLysGIuLysLeuGIuLysLeuArgLeuAlaAlaGIuInPheC 384
|||||
1113 CAAGAACAAAGAGAAAGCTAGAGAAAGCTGCCCTGGCGGCGCAGATTCT 1162
384 YrThrArgLeuGIyArgTyrArgMetProPheAlaI1eThrAlaValHis 400
|||||
1163 GCACCCGCTGGGGCGCTACCGCATGCCCTTCGCTGACGCGGCTGCAC 1212
401 LeuAlaAsnI1eValSerSerAlaGIyGIuLeuAspArgAspSerAspSe 417
|||||
1213 TTGGCAATCATGTGAGCAGCGCTGGGAGCTGACCGGAGCTGTGACTC 1262
417 rGIuGIyGIuArgArgProAlaI1eThrAspArgArgArgArgGIyProG 434
|||||
1263 GGAAGGCGAGCGCGCGCCAGCCTGACAGACCGCGCGCTGGGGGCCCC 1312
434 I1eAspArgAlaSerSerGIyAspAspAlaCysSerPheSerGIyPheArg 450
|||||
1313 AGGACCGGGCGAGTGTAGTGGGACGACGCGCTTCACCTTCTGCGCTCC 1362
451 ProAlaThrLeuThrValI1ThrAsnPhePheLysGIuGIuArgLe 467
|||||
1363 CCAGCCAGCGTACTGTACAAACTTCTTTAAGCAGAGAGGTGAGCCACT 1412
467 uSerAspGIuAspLeuPheLysPheLeuAlaAspMetArgArgProSerS 484
|||||
1413 CAGTACGAGGAGACCTCTCAAGTTCTCTGCTGACATGAGGCGCCGCTGT 1462
484 eTLeuLeuArgArgLeuArgProValI1ThrAlaGIuLeuLysI1eAspI1e 500

```
|||||
1463 CCCGTCGGCGGCGACTACGTCTCTGTACTGTCACAGCTCAAGATCGACATT 1512
501 SerProAlaProGluAsnProHisPheCysLeuSerProGluLeuLeuH 517
|||||
1513 TCCTCGGCTCTCTGAAAAATCCCACTTCTGTGCTCTCCCGTAGTGGCTCA 1562
517 sIleLeuProTyrProAspProArgGlyArgProThrLysGluIleLeuS 534
|||||
1563 TATCAAGCCCTTACCCTCGACCCGAGGGGCGCGCCACCAGAGAGATTCTGG 1612
534 IupheProAlaArgGluValTyrAlaProHisThrSerTyrArgAsnLeu 550
|||||
1613 AGTTCCCCCGCGCGAGAGTATAGCCCCCATACAGACTACAGAAACCTG 1662
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGlnLysEryA 567
|||||
1663 CTGTACGTGATCCCGCACAGCTTCACACTTCAGCAGCGCGGAGGGCTCCG 1712
567 IArgAsnLeuAlaValArgValGlnTyrMetThrGlyLysAspProSerG 584
1713 GCGCAACCTTGCTGTGTGGAGTGCATGACATGACAGGCGAGGACCCAGCC 1762
584 IAlaLeuProValIlePheGlyLysSerSerCysSerGluPheThrArg 600
|||||
1763 AGGCTCGCGGGTCACTTTGGCAAGTCCAGCTGCAGTGAATTTACCCGC 1812
601 GluAlaPheThrProValValTyrHisAsnLysSerProGluPheTyrG 617
1813 GAGGCTTCACACCGGTGTCTACATACATACAGTCCCGCGAGTTTACGA 1862
617 uGluPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuL 634
1863 GGAAGTTCAAGCTGATCTTCCAGCTTCGCGACAGAGAAACCATCACCTGC 1912
634 euPheThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeu 650
1913 TGTTCACTTCTACCAATGTCAGCTGCCAGCCCGCGGCGACTGCCCCTG 1962
651 GluThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArgLe 667
1963 GAGACACCCCTGGGCTTACTTGATGCCACTGCTGACAGACGGGGCCCT 2012
667 uArgThrGlyProPheCysLeuProValSerValAspGlnProProP 684
2013 GAGGACCGGCGCTTCTGTCTCCAGTGTCTGTGGACACGCCGCCGCCCA 2062
684 eTyrSerValLeuThrProAspValAlaLeuProGlyMetArgTrpVal 700
2063 GCTATTCCGTGCTCACACCGGATGTGGCGCTTCGGGCAATGCCCTGGGTG 2112
701 AspGlyHisLysGlyValPheSerValGluLeuThrAlaValSerEryA 717
2113 GACGGTCACAAAGGGCGTGTTCAGTGTGAGCTCAGACAGCCGTGCTCTGT 2162
717 IHisProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValL 734
2163 GCACCCCCCAGACCCCTACCTGGACAATTCCTCACCCGTGTGACAGTCC 2212
734 euGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 750
2213 TGGAGAGAGGAGCTTCCCATTCCTCGGCTCAAGGACACTGTGCTGACCGAG 2262
751 GlyAsnValGluGlnLeuLeuArgAlaSerLeuAlaAlaLeuArgLeuAl 767
2263 GGCACAGTGGAGCAAGAGCTGCGGCGCAGCTTGCGACACTGGCGCTGGC 2312
767 aserProGluProLeuValAlaPheSerHisHisValLeuAspLysLeuV 784
2313 CAGCCCCGAAACCCCTTGTCCTTCGCCACACAGCGTGGACAGAGCTCG 2362
784 aIArgLeuValIleArgProIleIleSerGlyGlnIleValAsnLeu 800
|||||
2363 TCCGTCGTGTCATCAGGCCCCCGATCATCAGTGGCCAGATTGTGAACCTG 2412
801 GlyArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHisAr 817
2413 GGCCTGGAGGAGCTTTGAAGCAATGGCCCATGTAGTACACCTTTGTACACG 2462
817 gSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaA 834
2463 GAGCTGGAGGAGCGCCAGAGATGCCGGGGTGCACATGCCACAGCTGGCTG 2512
834 IatTyrValHisTyrAlaPheArgGluLeuProGlyThrGluProSerLeuPro 850
2513 CCTACGTGCACATACCTTTTCCCTTCTCCTGTGCCTGACAGCCAGCCTCCG 2562
851 AspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArgLysE 867
2563 GATGGAGGCCCCCTCCAGTGCACAGTGCAGGCTGCACACTGAGCCCGGTGCT 2612
867 rGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerLysSerS 884
2613 TGGTGGCCCCGCAACCTTCATCTGGCGCGTTCACAAAGCATCAGCACCA 2662
884 ePAsnProAspLeuAlaValAlaIAProGlySerValAspAspGluValSer 900
2663 GCAACCTGTACTCCCGCTGGCGCTGCTGTGTGATGACGAGGTTTCC 2712
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
2713 CCATTCCTGGGCAGACAGCTGCTTCACGAGAGAGTGGCTGTGCAGTGGCT 2762
917 IValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934
2763 GGTTCAGACAGCATGGCGCTACGCGAGGCCATCTCCAGACACCCGTGGTCT 2812
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950
2813 TTTTCACTGCTATGTTGTAAGATATGCGCTGCACCTCTGCTTGTGGCAG 2862
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
2863 GCAGTAGACACACCCCGCAAGCTGGCTTCCCGGAGACCTTCCTGAGAGA 2912
967 rIleThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984
2913 CATCACTGCCTTGGTGGGCTGTGTGGGCTGAGGTCAACCCGTGTGCC 2962
984 IAspAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
2963 ACAAGAGATGGAGCTGGCCGAGCACTCAAGCCAGCCCTGGCTTTCTTC 3012
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
3013 CTCAGTGAACCTTCTGTCCCTGGTGGACCGGGGCTTGTGCTTCACACGTGT 3062
1017 IArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerProAsn 1034
3063 CCGGGCCCACTACAAGCAGGTGGCCACGCGGCTCCAGTCGTCCCTATATC 3112
1034 rAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCysSer 1050
3113 CAGCAGCCCTGTGACCTTGCGCATGGATTCACCCGATCTGTGAGC 3162
1051 HisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSerProPr 1067
3163 CACGAGCACTAGTGACCTCAACCTCCCTGCGTGGCCCGCTGTACCTCC 3212
1067 aIAserProSerProSerValSerSerThrThrSerGlnSerSerThrp 1084
3213 AGCTTCGGCTTCCTCTGTGTCTCCACACCACTCCACAGTGCACACT 3262
1084 heSerSerGlnAlaProAspProLysValThrSerMetPheGluLeuSer 1100
3263 TCTCCAGCCAAAGCCCGGAGCCCAAGGTGACCAAGATGTTGAACTGAGT 3312
```

1101 GlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLe 1117
|||||
3313 GGACCATTCGCGCAGCAGCAGCTCTAGCTGGCTCCTGCTGAGCGAGCT 3362
1117 uAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLysL 1134
|||||
3363 GGCACTGGCCCTCGAACCTGAGGCTGAAAGGGGCAATCCTGTTGCAAGA 3412
1134 ysaAlaIleSerAlaValHisSerLeuLeuGlyHisAspThrAspPro 1150
|||||
3413 AGGCCATCAGTCTGTGCACAGCCTGCTATGTGSCCATGACACGACCC 3462
1151 ArgThrAlaGluAlaThrValLysAlaArgValAlaGluLeuThrLeuPr 1167
|||||
3463 CGCTACGCGAGCGCACTGTGAAGGCTGTGTGCGCCAGCTGTACCTGCC 3512
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
|||||
3513 ACTGCTTCGATGTGACGGGATACCTTGGCCAGGCTGATGACTTTCTGTG 3562
1184 IuglyProGlyInArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
|||||
3563 AGGGCCCAAGGTACAGGGGTCAAGACTGGCTCAATGCTTGACTGACACA 3612
1201 GluGlyGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217
|||||
3613 GAAGCGCAAGGGGACATTGCGGGTACCATCAACCCCTGTGGCCATGGC 3662
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGln 1234
|||||
3663 CATTCGTGTGGCCCTCCTAGCCCTGTGGCTCCGGCCAGCATCTCCCAAG 3712
1234 IyProProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerSer 1250
|||||
3713 GGCCACCAACGGCTTCTGCGCAGCGTGTGCCCTCTCTGCTGATGATCAAG 3762
1251 ArgThrLeuLeuAlaCysValLeuThrValLeuLysAsnThrGluProAl 1267
|||||
3763 CGGACCTTGTGCGGTGTGTGCTGTGGTGTGTAACCAACCGGCGGGC 3812
1267 aLeuLeuGlnArgThrAlaThrAspLeuThrLeuProGluLeuGlyArgL 1284
|||||
3813 GCTCTGAGAGGCTGGGCGCACTGACCTGACACTCCCCAGCTGGGAGCTC 3862
1284 euleuAspLeuLeuThrLeuCysLeuAlaAlaPheGluThrLysGlyLys 1300
|||||
3863 TGTGGACTTGTCTGTACTTGTCTGCTGTGGCTTGTGATCAAGGGGAAA 3912
1301 LysAlaPheGluArgIleAsnSerLeuThrPheLysLysSerLeuAspMe 1317
|||||
3913 AAGGCTTTGAACGCAATCAACAGCTCACAATCAAAAATCTGTGGATAT 3962
1317 tLysAlaArgLeuGluGluAlaIleLeuGlyThrIleGlyAlaArgGln 1334
|||||
3963 GAAGCGCGGGCTAGAGGAAGCATTTGTGGTACCATCGAGCTGAGAACG 4012
1334 IuMetValArgArgSerArgGluArgSerProPheGlyAsnProGluAsn 1350
|||||
4013 AAATGGTTCCGGCAAGTCTGTAGAGAGCCGTTTGGGAATCCAGAGAAC 4062
1351 ValArgThrArgLysSerValThrHisThrLysGlnThrSerAspArgVa 1367
|||||
4063 GTGGCGTGGCGGAAGAGGTACACACTGGGAAGCAAACTGAGACCGCGT 4112
1367 IAspLysThrLysAspGluMetGluHisGluAlaLeuValGluGlyAsnL 1384
|||||
4113 GGACAAAGCAAGGATGAATGAACAACAGAGCCTTGTGGAAGGAAACC 4162
1384 euAlaThrGluAlaSerLeuValValLeuAspThrLeuGluIleLeuVal 1400
|||||
4163 TGGCAACCGAGCAAGCCTAGTCTGTCTGACACACTGAGATCATCTGTG 4212

1401 GlnThrValMetLeuSerGluAlaArgGluSerValLeuGlyAlaValLe 1417
|||||
4213 CAGACGGTATGCTTTCAGAAAGCCCGGAGAGCGTCTTGGGGCGAGTGT 4262
1417 uLysValValLeuThrYrSerLeuGlySerIleGlnSerAlaLeuPheLeuG 1434
|||||
4263 GAAGTTTGCTGTACAGCCTGGGCGAGTCCAGAGTGGCCCTTCTTCC 4312
1434 IuHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPheProGluLeu 1450
|||||
4313 ASCATGGCTGGGCGACCCAGAGGCCCTTGTGTCCAAAGTTCGCGAGCTG 4362
1451 LeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLeuArgLeuLe 1467
|||||
4363 CTGTTGAGAGAGACACGAGGCTGTGTGCCACCTGTGCTAGAGCTCTCT 4412
1467 uArgHisCysGlySerArgIleSerThrIleArgThrHisAlaSerAlas 1484
|||||
4413 ACGACACTGTGGACGGCGATACGACCATCCGCAAGCAGCCAGCGGCT 4462
1484 eIleuThrLeuLeuMetArgGlnAsnPheGluIleGlyHisAsnPheAla 1500
|||||
4463 CGCTATACCTGCTCATGCGACAGAACTTGAGATCGGCGCAACAATTTGCC 4512
1501 ArgValLysMetGlnValThrMetSerLeuSerSerLeuValGlyThrTh 1517
|||||
4513 CGGTGAAGATGCAAGGTCAACATGTCTCTGCTGCTGGTGGGAGCGAC 4562
1517 rGlnAsnPheSerGluGluHisIleuArgArgSerLeuLysThrIleLeuT 1534
|||||
4563 GCAGAACTTCAGTGAAGACACCTGGAGCGTTCACTCAAAAACCACTCTCA 4612
1534 hTrpAlaGluGluAspMetGlyLeuArgAspSerThrPheAlaGluGln 1550
|||||
4613 CCTATCTGAGGAGGACATGGGGCTGGGGAGACAGCACTTCCAGAGCAG 4662
1551 ValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLy 1567
|||||
4663 GTCCAGGACCTGATGTTCAACCTGCACATGATCTGCAGGACCGGTGA 4712
1567 sMetLysGluHisGlnGluAspProGluMetLeuIleAspLeuMetTYrA 1584
|||||
4713 GATGAAGAAACCAACAGAGACCTGAGATGCTCATCGACTCAATGTACA 4762
1584 rglAlaArgGlyTyrgingLysSerProAspLeuArgLeuThrTrpLeu 1600
|||||
4763 GAATTCCTCCGGGGCTACAGGGCTCACCGACCTTCGGCTGACCTGGTGG 4812
1601 GlnAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAl 1617
|||||
4813 CAGAACATGGCGGGAAGCAGCGGAGCTGGGCAACCAACGCGAGGCGGC 4862
1617 agLncYsMetValHisAlaAlaAlaLeuValAlaGluThrLeuAlaLeuL 1634
|||||
4863 CCAGTGCATGTGTGACAGCGCGCCCTGTGGCTGAGTACCTCGCCCTGCC 4912
1634 euGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIle 1650
|||||
4913 TCGAGAGCACGAGCCACCTGCGCGGTGGCTGTCTTCCAGAACATC 4962
1651 SerSerAsnValLeuGluGluSerAlaIleSerAspAspIleLeuSerPr 1667
|||||
4963 TCATCCAAAGCTGTAGAGAGATCCGCCATCTCCGAGACATCTCTGCC 5012
1667 oAspGluGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuV 1684
|||||
5013 CGAGCAGAGAGGGCTTCTGCTCGGGAAGCACTTCACTGAGCTGGGGCTGG 5062
1684 aIGlyLeuLeuGluGlnAlaAlaGlyTyrrPheThrMetGlyLysLeuTyrr 1700
|||||
5063 TAGGGTTGCTGGAAAGGCAAGCGCGGTACTTCAACATGGCGGGCTTAC 5112
1701 GluAlaValAsnGluValTyrrLysAsnLeuIleProIleLeuGluAlaHis 1717

```
|||||
5113 GAGGCGGTGAATGAGGTCTACAGAACCTCATCCCATCCGGAAGGCCA 5162
1777 SARGASPIRYLYSYLSLEUALAALVALHISGLYLSLEUGLSLUALAP 1734
|||||
5163 CCGTGACTACAAAGAGAGTGGCCGGGTGCACGGCAAACTGCAGAGAGGCT 5212
1734 hETHRLYSILMERHISGLINSERSERGLYTPGLUARGVALPHGGLYTHR 1750
|||||
5213 TCACCAATATCATCCACAGAGTTCCGGCTGGAGCGGCTGTTCGGGAGCG 5262
1751 TYRPHEARVALGLYPHETRYGLYALAHISPHGLYASPLEUASPGIUGL 1767
|||||
5263 TATTTCCGCGGTGGCTTCTACGGGGCCACCTTGGGTGACCTCGAATGAGCA 5312
1767 nGLUPHEVALYTRYLSGLUPROSERILETHRYSLEUALAGLULESERH 1784
|||||
5313 GGAGTTTGCTACACAGACCATCATCAGAACCTGCAGAGATCTCAC 5362
1784 ISARGLEUGLUGLUPHETRYTHRGLUARGPHEGLYASPARVALVALGLU 1800
|||||
5363 ACCGGCTGAGAGATTCTACACGAGAGATTGGCGACGAGCTGTGGAG 5412
1801 ILEILEYASPSERTYRPROVALASPLYSETRYLSLEUASPSERGLNLY 1817
|||||
5413 AFTATCAAGACTCTTACCCCTGTGGACAGTCCAAAGCTTGACTACAAAA 5462
1817 SALATYRILEGLINLETHRTRYVALGLUPROTYRPHASPTHYRGLUL 1834
|||||
5463 GGCTACTATCCAGATCAGTATGTGGAACCGTACTTGATACCTACGAGC 5512
1834 EULYSASPARVALYTHRYPHEASPARGASNTYRGLYLEUARGTYRPH 1850
|||||
5513 TCAAGAGACGGGTGACTACTTTACCGCACTATAGGGCTTCGCACTTC 5562
1851 LEUPHECYSTHRPROPHETHRPROASPLIYARGALAHISGLYLULEUPR 1867
|||||
5563 CTGTCTGTCACAGCGCTTCACCGCCGATGGGGCGCACACGGGAGGTGCC 5612
1867 OGUGLHISYSARGLYSTHRLLEULESERTHASPHISALAPHEROT 1884
|||||
5613 CGAGCAACACAAAGCGTAAAGCGTCTCAGACACGACGACCCCTTCCCT 5662
1884 YRILELSTHARCILEARVALYSHISARGIUGIUTHRVALLEUTHR 1900
|||||
5663 ACATCAAGACTCGCATCCGTGTGTGCCACCGGAGGAGCGTCTCAGC 5712
1901 PROVALGLUVALAILEGLUASPMETGLNLSYLSYTHRARGLULEUAL 1917
|||||
5713 CCAGTGGAGGTGGCCATCGAGGACATGCAGAAAGACAGGGAGGTGGC 5762
1917 APHEALATHRCGLUCINASPPROPROASPLIALYSMETLEUGLIMETVAL 1934
|||||
5763 CTTTGCCACCGAGAGAGACCCACAGATGCTAAGATGCTACAGATGTGC 5812
1934 EUGINGLYSERVALGLYPROTHRYVALASNGINGLYPROLEUGLIVALA 1950
|||||
5813 TTGAGGGCTGTGAGGGCCACCCCTGAACAGAGTCCCTCGAGAGTGGCC 5862
1951 GLINVALPHELEUALAGLUILEPROGLUASPPROLYSLEUPHEARGHISH 1967
|||||
5863 CAGGTGTTTTTAGCAGAGATCCCGAAGACCCCAAGCTTCCCGCATCA 5912
1967 SASNLYSLEUARGLEUCYSRPHLYSASPHNECYSLYLSYSCYSGLSUSPA 1984
|||||
5913 CAARAAATTCGGGCTGTGCTCAAGGACTTCTGAAAGAAATGTAGAGATG 5962
1984 IALEUARGLYSASNLYSALALEUULEGLYPROASPLINLYSGIUTYRHIS 2000
|||||
5963 CGCTGCGGAAAATTAAGGCCCTGATTTGGGCCGAGCCAGAAAGAGTACAC 6012
2001 ARGULIUEUGLUARGASNTYRCYSARLEUARGLUALALEUGLIPROLE 2017
|||||
```

```
6013 CGTAGCTGAGGCGCAACTACTCCCGCTGGCGGAGGCTCTGAGCCCT 6062
2017 uLEUTHRGILARGLEUPROGLINLEUMETALAPROTHRPROFGLYLEUA 2034
|||||
6063 GCTTACCAGCGCCCTGCCAGCTGATGGCACCCACCCCGGGCTCA 6112
2034 rGASNSERLEUASARGALASERPHARGLYSALASPLEU 2047
|||||
6113 GGAAGTCTTTGAACAGACGAAGTTTCCGAAAGGCAACACTC 6153
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/AA2001A.DAT:AAH41934
seq_documentation_block:
ID AAH41934 standard; cDNA: 6828 BP.
XX
XX AAH41934;
AC
XX
DT 30-AUG-2001 (first entry)
XX
DE Human CLASP-3 cDNA sequence SEQ ID NO:1.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
KW cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
KW antiinflammatory; antiarthritic; antianaemic; dermatological; uropathic;
KW ophthalmological; antirheumatic; nephroretic; antithyroid; antidiabetic;
KW neuroprotective; antiaclimatic; antibacterial; antisense therapy;
KW gene therapy; chromosome 1p31.1; ss.
XX
OS Homo sapiens.
XX
PN WO200142297-A2.
XX
PD 14-JUN-2001.
XX
PF 13-DEC-2000; 2000WO-US34171.
XX
PR 13-DEC-1999; 99US-0170453.
XX
PR 14-JAN-2000; 2000US-0176195.
XX
PR 14-FEB-2000; 2000US-0182296.
XX
PR 11-APR-2000; 2000US-0196267.
XX
PR 11-APR-2000; 2000US-0196460.
XX
PR 11-APR-2000; 2000US-0196527.
XX
PR 11-APR-2000; 2000US-0196528.
XX
PR 11-APR-2000; 2000US-05472276.
XX
PR 13-OCT-2000; 2000US-0240503.
XX
PR 13-OCT-2000; 2000US-0240508.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PI Lu P, Gartman JD, Candia AF;
XX
DR WPI: 2001-375003/39.
XX
DR P-PSDB: AAB99495.
XX
PT Novel isolated cadherin-like asymmetry protein (CLASP) -3, useful for
PT useful in preventing or treating a CLASP-3-mediated disease e.g.
PT autoimmune disease
XX
XX Claim 1; Fig 6; 189pp; English.
XX
XX The present invention describes an isolated polypeptide (I) comprising a
XX nucleotide (nt) sequence (S1a) that has at least 90% identity to the
XX sequence given in AAH41934 and is immunologically cross-reactive with
XX the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a
XX biological function with native CLASP-3. (I) has immunosuppressive,
XX antiinflammatory, antirheumatic, antianaemic, dermatological, uropathic,
XX ophthalmological, antirheumatic, nephroretic, antithyroid, antidiabetic,
XX neuroprotective, antiaclimatic and antibacterial activities, and can be
XX used in antisense therapy, vaccine production and gene therapy. CLASP-3
XX related sequences can be used in preventing or treating a CLASP-3
XX mediated disease, preferably an autoimmune disease by inhibiting an
XX immune response. The autoimmune disease is caused or exacerbated by
XX increased activity of T helper 1 lymphocytes (TH1). AAH41893 to AAH41953
```

CC and AAB99491 to AAB99507 represent sequences which are used in the
CC exemplification of the present invention. CLASP-3 is localised in the
CC chromosome location 1p31.1.

XX Sequence 6828 BP; 2082 A; 1384 C; 1432 G; 1930 T; 0 other;

alignment_scores:
Ratio: 7198.50 Length: 2140
Gaps: 3.960
Percent Similarity: 84.953 Percent Identity: 64.953

alignment_block:
US-09-736-968a-2 x AAH41934 ..

Align seg 1/1 to: AAH41934 from: 1 to: 6828

```
2 AAlaSerGluATGATGAlaPheAlaHisLysIleAsnArgThrValAl 18
||| ::::::::::::::::::::::::::::::::::::::::::::::
19 GCCATGGCCGCGCGCGCCCTTCCGCAAGATCAGCAAGCGGTGC 68
||| ::::::::::::::::::::::::::::::::::::::::::::::
18 aAlaGluValArgLysGlnValSerArgLuarGSerGlySerProHis 35
||| ::::::::::::::::::::::::::::::::::::::::::::::
69 AGCCGAATTAGGAAGCAGATCTCCGACAATATAGTGTCTCCCAAC 118
||| ::::::::::::::::::::::::::::::::::::::::::::::
35 eSerArg.....ArgCysSerSerSerLeuGlyVal 45
::: |||
119 TGCTCAAAACCTTAATATGTGTGCAATATATCCATCAACACACAGTG 168
||| ::::::::::::::::::::::::::::::::::::::::::::::
46 ProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeuSe 62
||| ::::::::::::::::::::::::::::::::::::::::::::::
169 CCCCTTACCGAAGCAGATGATCCAGATTTGGAGATTACCATTC 218
||| ::::::::::::::::::::::::::::::::::::::::::::::
62 rArgProAspAlaGluProGlyProLeuArgAspLeuValGluPhe 79
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
219 TCATCTCTTGGCTGTGATTTCTGGGCTTTACGGGATTTGATTAATTTC 268
||| ::::::::::::::::::::::::::::::::::::::::::::::
79 roAlaAspAspLeuGluLeuLeuLeuGlnProArgLuarGlySerThr 95
||| ::::::::::::::::::::::::::::::::::::::::::::::
269 CTCAGATGATATTGAAGTTGATTATGCTCTCGGACTCGAAGCTCT 318
||| ::::::::::::::::::::::::::::::::::::::::::::::
96 GluProGlyLeuProLysAspGluLysLeuAspAlaGlnValArgAla 112
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
319 GTTTCAGCTGTACCTGAAGAAAGTAAATGATCCACATGTTAAGACTG 368
||| ::::::::::::::::::::::::::::::::::::::::::::::
112 aValGluMetArgLysLeuAspTrpValIleValHisArgArgTrpGln 129
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
369 TATAAGAAATTATACAGAAAGCTGGCAATTTGTATCAGAAAATATCATA 418
||| ::::::::::::::::::::::::::::::::::::::::::::::
129 yrLeuSerAlaAlaLysSerProValThrThrAspThrGlnArgGluArg 145
||| ::::::::::::::::::::::::::::::::::::::::::::::
419 AATTGGGAACAGATTAAATCCCAATACATTAGATAAACAAGAAAGAG 468
||| ::::::::::::::::::::::::::::::::::::::::::::::
146 GlnLysGlyLeuProArgGlnValPheGluGlnAspAlaSerGlyAspG 162
||| ::::::::::::::::::::::::::::::::::::::::::::::
469 CAAAAGGTTTGGCCAAAACAAGTTTGAATCTGATGAAGCTCCAGATGG 518
||| ::::::::::::::::::::::::::::::::::::::::::::::
162 uArgSerGlyProGluAspSerAsnAspSerArgArgGlySerGlySerP 179
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
519 CAACAGCTACCAAGATGATCAAGATGACCTTAAAGACGTTCAATGTCAA 568
||| ::::::::::::::::::::::::::::::::::::::::::::::
179 roGluAspThrProArgSerSerGlyLysSerIlePheAspLeuArg 195
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
569 TAGATGATACCCCAAGGGGTAGCTGGGCTGTACTATCTTTGACTTCAA 618
||| ::::::::::::::::::::::::::::::::::::::::::::::
196 AsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaPr 212
||| ::::::::::::::::::::::::::::::::::::::::::::::
619 AATTCACCTCTCTGATGCTTGTCTTCCCAATTTACTGATGAACTCCAAA 668
||| ::::::::::::::::::::::::::::::::::::::::::::::
212 oGluAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgProp 229
||| ::::::::::::::::::::::::::::::::::::::::::::::
669 TGAAGAAATTAAGCCGTGAGATGATGACCAAGAAATCAAAACCGTCA 718
```

```
229 roAlaLeuLeuThrLeuArgProAlaProAspGluAspGluAlaValGlu 245
||| ::::::::::::::::::::::::::::::::::::::::::::::
719 AAGAACTTTTGTCTTGCATTCATCCACGAGATGAGGAAGAAACCATTA 768
||| ::::::::::::::::::::::::::::::::::::::::::::::
246 ArgCysSerArgProGluProArgGlnHisPheGlyGlnArgIleLe 262
||| ::::::::::::::::::::::::::::::::::::::::::::::
769 CGGCTTAGTGTCTCGATATACCAAGAAACATTTTGGTCAAGACTTCT 818
||| ::::::::::::::::::::::::::::::::::::::::::::::
262 uValLysCysLeuSerLeuLysPheGluLysLeuLeuProIlePheG 279
||| ::::::::::::::::::::::::::::::::::::::::::::::
819 TGTAAATCTTATCATCTCAAGTTGAAATTTGAATTCAGACCATTTTGT 868
||| ::::::::::::::::::::::::::::::::::::::::::::::
279 yLLeuAlaLeuArgAspValArgGluLysLysLysIleSerGluAsn 295
::: ||| ::::::::::::::::::::::::::::::::::::::::::::::
869 CAAGTTTGGCTTTATATGATGTCAAGAAAGAAAGAAATTTCAAGAAAC 918
||| ::::::::::::::::::::::::::::::::::::::::::::::
296 PheTyrrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 312
||| ::::::::::::::::::::::::::::::::::::::::::::::
919 TTTTATTGTGACCTTATATCTGACGAGATGAAAGGTTGTTCGTCCACA 968
||| ::::::::::::::::::::::::::::::::::::::::::::::
312 sGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSer 329
||| ::::::::::::::::::::::::::::::::::::::::::::::
969 TGTACACACTGCTGCATTAATCACTCCCTGGCAGATCAGCAATTTTCTA 1018
||| ::::::::::::::::::::::::::::::::::::::::::::::
329 aThrTyrrProSerProAspIlePheLeuValIleLysLeuGluLysVal 345
||| ::::::::::::::::::::::::::::::::::::::::::::::
1019 TCACCTATCTCTCCCAAGATGTTTCTTGTATTAAGCTAAGAAAGTTC 1068
||| ::::::::::::::::::::::::::::::::::::::::::::::
346 LeuGlnGlnGlyAspIleSerGlyCysGluProTyrrMetValLeuL 362
||| ::::::::::::::::::::::::::::::::::::::::::::::
1069 CTACACCAAGAGACATTGAGAGATGTGCAGAACCATATATGATTTTCAA 1118
||| ::::::::::::::::::::::::::::::::::::::::::::::
362 sGluValAspThrAlaLysAsnLysGluLysLeuGluLysLeuArgLeu 379
||| ::::::::::::::::::::::::::::::::::::::::::::::
1119 AGAAGCAGATGCGCACCAAGATTAAGAAACTGGAGAACTGAGAGAGTC 1168
||| ::::::::::::::::::::::::::::::::::::::::::::::
379 laAlaGluGlnPheCysThrArgLeuGlyArgTyrrArgMetProPheAla 395
||| ::::::::::::::::::::::::::::::::::::::::::::::
1169 AAGCAGATCAGTTTGTCCCAAGACTGGGCAATATCGCTTGTCTTTCCT 1218
||| ::::::::::::::::::::::::::::::::::::::::::::::
396 TrpThrAlaValHisLeuAlaAsnIleValSerSerAlaGlnGluLeuAs 412
||| ::::::::::::::::::::::::::::::::::::::::::::::
1219 TGGACTGCAATCCATTTAATGAATATTTGTTACAGATGCTGGAGAGTTTGA 1268
||| ::::::::::::::::::::::::::::::::::::::::::::::
412 pArgAspSerAsp.....SerGluGlyLuarArgArgProAlaT 425
||| ::::::::::::::::::::::::::::::::::::::::::::::
1269 AAGAGTTTCTACAGAAAGTAAATCAGTACTGAGACAGAAAGGCTCTT 1318
||| ::::::::::::::::::::::::::::::::::::::::::::::
425 rPThrAspArgArg.....ArgGlyProGlnAspArg 436
||| ::::::::::::::::::::::::::::::::::::::::::::::
1319 GGTCAAGAGAGAGAAATTCTAGTATTTGTGCGACGATCACTTGAAGG 1368
||| ::::::::::::::::::::::::::::::::::::::::::::::
437 AlaSerSerGlyAspAspAlaCysSerPheSerGlyPheArgProAlaTh 453
||| ::::::::::::::::::::::::::::::::::::::::::::::
1369 ACACACAGGAGAGATGATCTGTGAATCTTGAGAGAGCTTTCGACAGCTAC 1418
||| ::::::::::::::::::::::::::::::::::::::::::::::
453 rLeuThrValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspG 470
||| ::::::::::::::::::::::::::::::::::::::::::::::
1419 TGTCAAGTACCAAAATTTTAAAGCAGAGAAAGACCGCTTAATGATG 1468
||| ::::::::::::::::::::::::::::::::::::::::::::::
470 LuAspLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeu 486
||| ::::::::::::::::::::::::::::::::::::::::::::::
1469 AAGATCTTACAAATCTCTGCTGATATGAGAAGGCCATCTTCTGTCTTA 1518
||| ::::::::::::::::::::::::::::::::::::::::::::::
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
||| ::::::::::::::::::::::::::::::::::::::::::::::
1519 CGGCGCTAAGACCTTAATACACTCAGCTCAAGATGACATTTCCTCCGC 1568
||| ::::::::::::::::::::::::::::::::::::::::::::::
503 aProGluAsnProHisPheCysLeuSerProGluLeuLeuHisLysLysP 520
||| ::::::::::::::::::::::::::::::::::::::::::::::
1569 ACCTGAAATTCGCCATATATGCTTAAGTCCGAGAGCTGCTTCAAGTAAAG 1618
||| ::::::::::::::::::::::::::::::::::::::::::::::
520 rOTyrrProAspProArgLysArgProThrLysGluLysLeuLeuPhePro 536
```

```
||||| ||| |||||||:|||||
1619 TTACCCCTCAGTAGAGTTACAGCTACAGAAATCTTAGAGTTTCCC 1668
537 AlAaRGlValTfAlProhIshSerTyARgaSnLeuTYVa 553
|||||:|||||:|||||:|||||:|||||:
1669 GCAGGGAGTTTATGTTCAAACACTACTACAAATCTTCTCTACAT 1718
553 lYrProhISerLeuAaSnPheSerSerArgGlnGlySerValARgaSnL 570
|||||:|||||:|||||:|||||:|||||:
1719 ATACCCCTCAGACTTAAATTTTGCACATCGTCAAGTCTCTAGAAATA 1768
570 euAlaValARgaSnLInTYrMetHrGlnGluAaSProSerGlnAlaLeu 586
|||||:|||||:|||||:|||||:|||||:
1769 TACACGTGAAGTCCAGTTTATGTATGAGAGATCCAGCAATGCCATG 1818
587 ProValIlePheGlyLysSerSerCysSerGlnPheHrArgGlnAlaPh 603
|||||:|||||:|||||:|||||:|||||:
1819 CCGGTAACTTTGGTAAATCTAGCTGTTCAGAAATTTTCAAGGAAGCTTA 1868
603 eHrProValValTYrHISaSnLysSerProGlnPheTYrGlnGluPheL 620
|||||:|||||:|||||:|||||:|||||:
1869 TACAGCCGTATATCATACAGAGCTCTCTCATTTTCATGAAAGAAATCA 1918
620 ySLeuHISLeuProAlaCysValHrGlnAaSnHISHISLeuPheHr 636
|||||:|||||:|||||:|||||:|||||:
1919 AGGTTAAGCTCTCTGCTACTTTAACTGACATCATCTGCTTTTACT 1968
637 PheTYrHISValSerCysGlnProARgProGlyHrAlaLeuGlnTYrPr 653
|||||:|||||:|||||:|||||:|||||:
1969 TTTTATCATGTATAGTGTACAAACAAATACTCCTCTTGAAACACC 2018
653 oValGlyPheHrTYrPleProLeuGlnHISGlyARgLeuARgTYrG 670
|||||:|||||:|||||:|||||:|||||:
2019 AGTTGATATACATGATACCATGCTTCCAGAAATGAGAGTTGAAGACTG 2068
670 lYrPheCysLeuProValSerValaSPGlnProProSerSerTYrSer 686
|||||:|||||:|||||:|||||:|||||:
2069 GCCAGTTTGTCCAGTCTCATTTGGAAACCCACAGGCTTATTTCT 2118
687 ValLeuTYrProAspValAlaLeuProGlyMetA-gTYrValaSPGlyNI 703
|||||:|||||:|||||:|||||:|||||:
2119 GTACTGTCTCGTAGGCTCTCACTGCGCATGAATGAGTATATATCA 2168
703 sLYSGlyValaPheSerValGlnLeuTYrAlaValSerSerValHISProG 720
|||||:|||||:|||||:|||||:|||||:
2169 CAAGAGTGTATATGTTGAAGTGTCTGTTCGTCTATCCATACAC 2218
720 lAaSProTYrLeuAspLYSphPheHrLeuValHISValLeuGlnGlu 736
|||||:|||||:|||||:|||||:|||||:
2219 AAGATCTTATCTTGACAAATTTTGTCTGTCATGCTCTGGATGAA 2268
737 GlyAlaPheProPheARgLeuLYSsPThrValLeuSerGlnGlyAspTY 753
|||||:|||||:|||||:|||||:|||||:
2269 CACCTGTCTCCAGTCCGAATTTGGGACATGCGAATCATGGAAATAACTT 2318
753 lGlnGlnGluLeuARgAlaSerLeuAlaAlaLeuARgLeuAlaSerProG 770
|||||:|||||:|||||:|||||:|||||:
2319 AGAAATGAAATGAAGACAGCATTTTACAGACTGAATCATCCACGCTGG 2368
770 lUProLeuValAlaPheSerHISHISValLeuAspLYSLeuValARgLeu 786
|||||:|||||:|||||:|||||:|||||:
2369 AACCAATGAGTCCGATTTCTTCAATTTCTGCTAGATTAACGTACTTTTAA 2418
787 ValIleArgProProIleSerGlnIleValaSnLeuGlyARgI 803
|||||:|||||:|||||:|||||:|||||:
2419 GTATTTAGACCTCTCTCATCTGCTGCCAAATAGTTAACTAGTGTCAACG 2468
803 yAlaPheGlnAlaMetAlaHISValaSerLeuValHISArgSerLeuG 820
|||||:|||||:|||||:|||||:|||||:
2469 ATCTTTTGAAGCATAGCATCAATTTAAATGCACTTCACAAAACCTTGG 2518
820 lUAlaAlaGlnAspAlaArgGlyHISCysProGlnLeuAlaAlaTYrAl 836
|||||:|||||:|||||:|||||:|||||:
2519 AAGAAATCATGACCAAGATGGCAGAAACAGCCTTCTTGATCATATATT 2568
837 HlSTYrAlaPheARgLeuProGlyTYrGlnUProSerLeuProAspGlyAl 853
|||||:|||||:|||||:|||||:|||||:
2569 CATTTATGTTTCCGCTACCAAAATACTTACCTTATATCATCATCACAGG 2618
853 aPro.....ProValTYrValGlnAlaAlaThLeuAlaArgGlyS 867
|||||:|||||:|||||:|||||:|||||:
2619 TCCGTGGGGTTTGGAGAGATCATGCTCATATGCCACAAATGCTGATCTG 2668
867 eGlyARgProAlaSerLeuTYrLeuAlaArgSerLYSserIleSerSer 883
|||||:|||||:|||||:|||||:|||||:
2669 CCGTGAAGACCTGCAGCCTTATTTAAATCGTTCTCGAAGCTTAGTAAT 2718
884 SerAsnProAspLeuAlaValAlaPheGlySerValaSPasGlyValaSe 900
|||||:|||||:|||||:|||||:|||||:
2719 AGCAATCCAGATATATCTGGGACTCCACGTCACAGATGAAATTCG 2768
900 rArgIleuAlaSer..... 905
|||||:|||||:
2769 ATCAATCATCGGAGATGAAGGTTTAGATCGCTCCCAATTCCTGGTTACA 2818
905 ..... 905
2819 CTGGTGTCCAAAGCTGCCCATGGGAGTCCAAACCCAGTCCAAATGCA 2868
905 ..... 905
2869 GATTCACACAGGCTATGATGATGAGTGTATGATGTCTTGCAAC 2918
906 .....L 906
2919 AGAGACGTCAAGTTTCTTACAAACATTAAACGGAGCCTTACCAACTAAA 2968
906 ySLeuGlnHISGlnGluLeuAlaLeuGlnTYrValaValSerSerSerAla 922
|||||:|||||:|||||:|||||:|||||:
2969 AGCTTTTACAGAGAGCTGAGCTTTCAGTGGGTGTGTGAGAGGGACAGC 3018
923 ValArgGlnAlaIleLeuGlnHISAlaTYrPhePhePheGlnLeuMetVa 939
|||||:|||||:|||||:|||||:|||||:
3019 GTTCGGGAATCAGCTTTGCACACAGCCTTGCTTTTGTGAATTAATAGT 3068
939 lLYSerMetAlaLeuHISLeuLeuLeuGlyGlnARgLeuAspTYrProA 956
|||||:|||||:|||||:|||||:|||||:
3069 AAAGACATAGTGCACACCATTTTACTTATATGATATAACTTAGAGCTCAA 3118
956 rGlySLeuARgPheProGlyARgPheLeuAspAspIleHrAlaLeuVal 972
|||||:|||||:|||||:|||||:|||||:
3119 GGAAGAGTGTGTTTCCAGAACGTTTCATGATGATGACATGCAAGCTTGTG 3168
973 GlySerValGlyLeuGlnValIleHrArgValHISLYSAspValGlnLe 989
|||||:|||||:|||||:|||||:|||||:
3169 AGCAGCATTTGCTAGTATATAGTTTACAGATTTTCAGAAAGACACAGAAAT 3218
989 uAlaGlnHISLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuS 1006
|||||:|||||:|||||:|||||:|||||:
3219 GCTTGAGAGACATCAATACAAACCTTGCAATCTTCTCATATATCTGTTGT 3268
1006 eTYrLeuValaSPasrGlyPheValaPheSerLeuValaArgAlaHISTYrLYS 1022
|||||:|||||:|||||:|||||:|||||:
3269 CTGTATGACACAGAGATTTGTTTATGCTTATTAAGCTTCGCTATTA 3318
3269 CTGTATGACACAGAGATTTGTTTATGCTTATTAAGCTTCGCTATTA 3318
|||||:|||||:|||||:|||||:|||||:
3319 GATTAAGCTTACCGCTCAGCTTACTTACTTCCACCTGCATCTCCATCACT 3468
1039 rLeuARgMetGlnPheHrArgIleLeuCysSerHISGlnHISTYrValIT 1056
|||||:|||||:|||||:|||||:|||||:
3369 CTTGAGGCTGGATTTTCTAGCAATCATCTGCACTATAGACATATGTGA 3418
1056 hTYrLeuAsnLeuProCysCysProLeuSerProProAlaSerProSerPro 1072
|||||:|||||:|||||:|||||:|||||:
3419 CATTAACCTTACCGCTCAGCTTACTTACTTCCACCTGCATCTCCATCACT 3468
```



```
1073 SerValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaIpr 1089
|||||
1089 OAsProIysValThrSerMetPheGluLeuSerGlyProPheArgGln 1106
|||||
3519 AGACCAAAAGATTGGCAAAATATCTTGAATTTATCCCTGCTTCCGCCAAC 3568
|||||
1106 LnhISpHeuAlaGluLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlu 1122
|||||
3569 ACCATTATTGGCAGGACTTGTGTAAACAGAGCTGGCTGCTATTPTTAAC 3618
|||||
1123 ProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSerAlaVa 1139
|||||
3619 CCTGATGCTGAAGGACTGTTGGATTGCATTAAGAAATCATCAATCATGCT 3668
|||||
1139 LhISerLeuLeuCysGlyLhISpThrAspProArgTyraIaGluAlaIpr 1156
|||||
3669 ACACAAATTACTCTCCAGTCAGACTCAGACCCGGGTACTCTGCACCTC 3718
|||||
1156 hVAlLysAlaArgValAlaGluLeuTyrrLeuProLeuLeuSerIleAla 1172
|||||
3719 AGATAAGGCTCGAGGCGCATGTGTATCTACTCTGATTGGATTATATC 3768
|||||
1173 ArgAspThrLeuProArgLeuHisAspPheAlaGluGlyProGlyGlnAr 1189
|||||
3769 ATGGAAACGTACTCTAGCTGTATGATTTTACAGAACTCACAAATCAACG 3818
|||||
1189 gSerArgLysAlaSerMetLeuAspSerAspThrGluGluGluAspI 1206
|||||
3819 AGCAACACCAATTTGTATAGCCATGATGATTAAGTAGAGC.... 3864
|||||
1206 LeAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyPro 1222
|||||
3865 ..GGAAGTATGATTAACCCAGACCGTGGCCATGCGCAATCCAGGAGCATCG 3912
|||||
1223 Leu.....AlaProGlySerArgAlaSerIleSerGluGlyPr 1235
|||||
3913 GTCCCTCACTAACAGCGCTGGCACTTCCCTCAGCTCA..... 3954
|||||
1235 oProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGluSerSerArgT 1252
|||||
3955 ....ACGAGTGGCAGCAACACACTACTTTCAGCAGATCAAGTCGAA 4000
|||||
1252 hTrLeuAlaCysValLeuTrpValLeuLysAsnThrGluProAlaLeu 1268
|||||
4001 GCCTTTGATCTGTCTACTTGGTCTCAAAATGCAAGTCAACAGATT 4050
|||||
1269 LeuGlnArgTrpAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLe 1285
|||||
4051 CTACAGAGAGGTTTACAGATCTCTCAGTCTTGCACTTAACCGGCTATT 4100
|||||
1285 uAspLeuLeuTyrrLeuCysLeuAlaAlaPheGluTyrrLysGlyLysLysA 1302
|||||
4101 AGATCTGCTTATCTCTGTGTCTGCTTGTGAGTAAAGGAAAAAAG 4150
|||||
1302 LAbheGluArgGlyIleAsnSerLeuThrPheLysLysSerLeuAspMetLys 1318
|||||
4151 TGTGTGAACAATGATAGCTTACCTTTTAAGAAATCAAAAGCATGAGACA 4200
|||||
1319 AlaArgLeuGluGluAlaIleLeuGlyThrIleGlyAlaArgGlnLys 1335
|||||
4201 GCAAAGCTTGAAGAGCTATTCTGGAGCATAGTGCCAGCAAGAAAT 4250
|||||
1335 tValArgArgSerArg.....GluArgSerPro.....P 1345
|||||
4251 GGTACGGCGAAGCCGAGACAGCTCGAGAGAAACCATCTGGAAGTGCCT 4300
|||||
1345 hGluAsnProGluAsnValArgTrpArgLysSerValThrHisTrpLys 1361
|||||
4301 TTGGAACTCAAGAAATTTGAGGTGGAGGAAGAAATATGACTCTGCGCT 4350
|||||

1362 GlnThrSerAspArgValAspLysThrLysAspGluMetGlnHisGluAl 1378
|||||
4351 CAAAACACAGAGAAGCTGTGCAAAATCAAGAGCAGAGATTGAACACGAAAC 4400
|||||
1378 aLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuValValLeuAspT 1395
|||||
4401 ACTGATTGATGAGAAACCTGGCTACAGAGCAACCAATCAATCATTTAGATA 4450
|||||
1395 hTrLeuGluIleIleValGlnThrValMetLeuSerGlnAlaArgGlnSer 1411
|||||
4451 CATTAAGATTGTGTGTTCCAGACCGCTTCTGTACAGGAATCCAAAGAGAC 4500
|||||
1412 ValLeuGlyAlaValLeuLysValValLeuTyrrSerLeuGlySerAlaG 1428
|||||
4501 ATTCTGTGTGAGAGCTAAAGTGTACTACACAGATGGCCGTGAACCA 4550
|||||
1428 nSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValS 1445
|||||
4551 AAGTCAGTTTATCTACACACGTTTGTCTACACAGAGAGCCTTGTTT 4600
|||||
1445 eTrLysPheProGluLeuLeuPheGluGluLysAspThrGluLeuCysAlaAsp 1461
|||||
4601 CAAGTTTCTCGAAGCTCTTATTGTAACAGACAGACAGAGCTGTGCTGAT 4650
|||||
1462 LeuCysLeuArgLeuLeuAlaArgHisCysGlySerArgIleSerThrIleAr 1478
|||||
4651 TTATGCTCAGAGCTTCTCCAGACACTGACAGATGATCGGTCAATATACG 4700
|||||
1478 gThrHisAlaSerAlaSerLeuTyrrLeuLeuMetArgGlnAspPheGlu 1495
|||||
4701 GTCACACCCCGAGTCCCTCCCTTACTACTAATGAGCCAAACTTTGACA 4750
|||||
1495 LeGlnHisAsnPheAlaArgValLysMetGlnValThrMetSerLeuSer 1511
|||||
4751 TTGGGAATTAACCTTGGCCAGGTTAAATGCAGGTACCAATGTCACATATC 4800
|||||
1512 SerLeuValGlyThrThrGlnAsnPheSerGluGlnHisLeuArgArgSe 1528
|||||
4801 TCTTGTGGTGGACATCTCAGAAATTTTAATGAATAATCTTAAGACCTTC 4850
|||||
1528 rLeuLysThrIleLeuThrTyrrAlaGluGluAspMetGlyLeuArgAspS 1545
|||||
4851 TCTAAAGCTATATTGACATATGCTGAAGAAAGATCTGGAATGGGGAAA 4900
|||||
1545 eTrThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuHisMetIle 1561
|||||
4901 CAACATTCCTGATCAGATCAGGTCGAGATCTGTTTCAATCTCCAAATGATT 4950
|||||
1562 LeuThrAspThrValLysMetLysGlnHisGlnGluAspProGluMetLe 1578
|||||
4951 CTTCTGTACTGTGAATATGAAGACACACCAAGAGATCTGTAATGTT 5000
|||||
1578 uLLeAspLeuMetTyrrArgIleAlaArgGlyTyrrGlnGlySerProAspL 1595
|||||
5001 GATTGATATTAATGTACAAATTTGCCAAGGTTTACCAGACCTCTCCAGAG 5049
|||||
1595 euArgLeuThrTrpLeuGlnAsnMetAlaGlyLysHisAlaGluLeuGly 1611
|||||
5050 ..CGATTGACCTGCTTGCAGAACATGCGCAGGCAAGACATCGAAGCAAGC 5097
|||||
1612 AsnHisAlaGluAlaAlaGlnCysMetValHisAlaIleAlaLeuValAl 1628
|||||
5098 AATCATAGCTGAAGCTGCACATGCTCTAGTCCACATCAGACAGACATTGTC 5147
|||||
1628 aGluTyrrLeuAlaLeuLeuGluAspGlnArgHisLeuProValGlyCysV 1645
|||||
5148 TGAATATTGTAGCATGCTGAGAGACCGGAATATCTTCTGTGGGATGTG 5197
|||||
1645 aLSerPheGlnAsnIleSerSerAsnValLeuGluGluLysSerAlaIleSer 1661
|||||
5198 TAACATTTCAGAATATTTCAATCTAATGTGTTTAAGAAATTCGCGCTCA 5247
|||||
1662 AspAspIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisAsp 1678
|||||
```

```

|||||.....|
5248 GATGATGGGATATCTCCAGATGAAAGAGTATCTGCTCTGGAATAATCTT 5297
1678 eThGluLeuGlyLeuValGlyLeuLeuGluGluAlaAlaGlyTyrPhe 1695
5298 TACTGAGTCAGACTGTGGATTAAGTGAACAAGACAGCTGCTCTCT 5347
1695 hTmetGlyGlyLeuTyrGluAlaValAsnGluValTyrLysAsnLeuIle 1711
5348 CTATGGCTGCATGTATGAACAGTATGAAAGTTTACAAAGTCTTATT 5397
1712 ProIleuGluAlaHisArgAspTyrLysLysLeuAlaValHisGly 1728
5398 CCTATTCATGAGCTATATCGGATGCAAGAAGACTATCCCAATTCATGG 5447
1728 yLysLeuGluGluAlaPheThrLysIleMetHisGlnSerSerGlyTrp 1745
5448 TAACTTCAGAAAGCATTCAGCAAAATTTGTCATCAGATCAGTGGCTGG 5497
1745 LuArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAlaHisPhe 1761
5498 AGCGGATGTTGGCACTATTTGCTGTGTGTTTATGGAACCAAGTTC 5547
1762 GLyAspLeuAspGluGluGluPheValTyrLysGluProSerIleThrLy 1778
5548 GGGGATTTGGATGAAACAAGATTGTTTACAGAGCGCTGCATTAACCA 5597
1778 sLeuAlaGluIleSerHisArgLeuGluGluPheTyrThrGluArgPheG 1795
5598 ACTTGCAAGATATCTCAGATTGGAGGATTTTACGGAAGAAAGATTGG 5647
1795 LyAspAspValAlaGluIleIleLysAspSerTyrProValAspLysSer 1811
5648 GAGAGGATGTGGTGAAGTATCAAAAGCTCTAATCTGTAGACAAAGTGT 5697
1812 LysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyrValGluProTy 1828
5698 AAATTAATCTCTTAAACAAGCATATTCAGATTACATATGAGGACATA 5747
1828 rPheAspThrTyrGluLeuLysAspArgValThrTyrPheAspArgAsn 1845
5748 CTTTGACACATATGAGATGAGAGCAAGATCACCATTATTCGACAAAAAT 5797
1845 yrgLysLeuArgThrPheLeuPheCysThrProPheThrProAspGlyArg 1861
5798 ACAATCTTCTCGATTCATGTAACACCTTTACTTACTTACTGACCG 5847
1862 AlaHisGlyLysLeuProGluGlnHisLysArgLysThrLeuLeuSerTh 1878
5848 GCCCATGGGGAAGCTTCATGAACAATTCAAAAGAGAACCATCTGACTAC 5897
1878 rAspHisAlaPheProTyrIleLysThrArgIleArgValCysHisArg 1895
5898 GTCTCATGCTTCTCTTATATTAAAAAGGATGCAATGTCATCATTAAG 5947
1895 LuGluThrValLeuThrProValGluValAlaIleGluAspMetGlnLys 1911
5948 AAGAGATCATCTTAACACCAATGAAGTTGCTATTGAGGACATGAGAAA 5997
1912 LysThrArgGluLeuAlaPheAlaThrGluGlnAspProProAspAlaLy 1928
5998 AAGACACAGAGAGTTGSCATTTGCACACATCAGAGATCCGAGAACCCCA 6047
1928 sMetLeuGlnMetValLeuGluGlnLysSerValGlyProThrValAsnGln 1945
6048 AATGCTTCAGATGTACTCCAGGATCTGAGCACCAAGATGAATCAGG 6097
1945 LyProLeuGluValAlaGlnValPheLeuAlaGluIleProGluAspPro 1961
6098 GGCCCTTGGAAAGTTGCCAGGTTTCTGTGAATATCAGTGAACCA 6147
1962 LysLeuPheArgHisHisLysLysLeuArgLeuCysPheLysAspPheCy 1978
|||||.....|

```

```

6148 AACCTCTCAGACATCATATAAATACTGCAGCTCTGCTTTAAAGATTTC 6197
1978 sLysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1995
6198 TAAAGAGTGTAGATGCTTTAAGAAAATAAGAGATTAATTTGGCCGG 6247
1995 sPGLysGluGlyThrHisArgGluLeuGluArgAsnTyrCysArgLeuAr 2011
6248 TTCAAAGAGATTCAAAGGAATTC...GGGAAGTATCTTCCGCTTAA 6294
2011 gGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGlnLeuMetAla 2027
6295 AGAGGCCCTACAGCCCT...AGATCACAGAAAGTCCCTCAGTTATCCAAG 6341
2028 ProThr.....ProProGlyLeuArgAsnSerLeuAsnArgAlaSerPh 2042
6342 CCAGTTCCTGCTGTCCCTCCGACAGAGATTCCTTCAGTCGATAGCTT 6391
2042 eArgLysAlaAspLeu 2047
6392 TCGCAAAATGAGATCTC 6407

```

seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA507373

seq_documentation_block:

ID AA507373 standard; cDNA; 7215 BP.

AA507373:

26-SEP-2001 (first entry)

Human cDNA encoding CLASP-5.

Human; CLASP-5; cadherin-like asymmetry protein; immune gateway; immunogen; antibody; autoimmune disease; rheumatoid arthritis; multiple sclerosis; leukaemia; insulin dependent diabetes mellitus; acquired immunodeficiency syndrome; AIDS; ss.

Homo sapiens.

Location/Qualifiers

112..6048

/*tag- a

/product= "CLASP-5"

1498..1519

/*tag- b

/note= "PCR primer HC5gs1"

3087..3113

/*tag- c

/note= "Antisense oligonucleotide AA507384"

3087..3113

/*tag- d

/note= "Antisense oligonucleotide AA507384"

3232..3259

/*tag- e

/note= "Antisense oligonucleotide AA507384"

complement (3642..3660)

/*tag- f

/note= "PCR primer HC5As10b"

4728..4756

/*tag- g

/note= "Antisense oligonucleotide AA507384"

W0200142296-A2.

14-JUN-2001.

13-DEC-2000; 2000MO-US34163.

13-DEC-1999; 99US-0170453.

14-JAN-2000; 2000US-0176195.

14-FEB-2000; 2000US-0182296.

11-APR-2000; 2000US-0196267.


```

411 euaspargaspseraspser.....GluGlyGluArg 421
1166 AGCTAACTGATGTGGACTGTGCTGGGAGAAAGCCAGTGGGTGACGG 1215
422 ArgProAlaThrPheHisPargArgArgGlyProGlnAspArgAla 438
1216 AGGACATTGGCCCATCTAGAAAGCTT.....TCTGAAGAGACCCCT 1256
438 rSerGlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeu 455
1257 CTCCTGGAGGAAATGGGCTTGATCAACTTCAAAACCTCCACTCTGA 1306
455 hrValThrAsnPhenPheGlnGluAlaGluArgLeuSerAspGluAsp 471
1307 GCGTTAGCACTTTTTCACAGCAAGAAAGAGACCTTACGGATGAAGAC 1356
472 LeuPheLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArg 488
1357 TTATTTCAGTTTAACTGACTACAAAAGATCATCATCTTACAGAGAGC 1406
488 GLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPro 505
1407 AGTCAACTCAATTCAGGCTTGCTAGAGACTGAGATTCTACAGCTCAG 1456
505 LuAsnProHisPheCysLeuSerProGlnLeuLeuHisIleLysProTyr 521
1457 AGATCATCAATTCGTGTGACTCTGAATTCCTGCCCGTGAAGACCTTT 1506
522 ProAspProArgGlyArgProThrLysGluIleLeuGlnPheProAla 538
1507 CTGAAAGACGGGACAGCCCGCACAAAGATTTTGGAAATTTCCAAACAG 1556
538 gGluValTyrAlaPheHisThrSerTyrArgAsnLeuLeuTyrValTyr 555
1557 AGAAGTATATGCTCCACACTGTGACAGAAACCTCTATGCTTACC 1606
555 roHisSerLeuAsrPheSerSerArgGlnGlySerValArgAsnLeuAla 571
1607 CACAGAGGCTGAACCTTTGTAAACAAACTACATACAGCCGGAAACATTAC 1656
572 ValArgValGlnTyrMetThrGlnGlyAspProSerGlnAlaLeuPro 588
1657 ATAAAGATTCAGTTTATGTGTGAGAAAGATCGTAACATCGATGCCGGT 1706
588 lIlePheGlyLysSerSerCysSerGlnPheThrArgGluAlaPheThr 605
1707 CACTTTGGAAATCCAGCGGCGCTGATTTCTGAGAAAGTGTACACAG 1756
605 roValValTyrHisAsnLysSerProGlnPheTyrGlnGluPheLysLeu 621
1757 CTGTTACATACCAATTAAGTCTCTGACTTTATGAAAGAGTGAATAAT 1806
622 HisLeuProAlaCysValThrGluAsnHisIleLeuPheThrPheTyr 638
1807 AAGCTCCCGCTAAGCTACAGTAATACACCACTCTGTTCACCTTCA 1856
638 rHisValSerCysGlnProArgProGlyThrAlaLeuGlnTyrProValG 655
1857 CCATATCAGCTGTACAGCAAGCAAGAGAGCCCTCGTGAACCTCTCTGG 1906
655 lYpHeThrTrpIleProLeuLeuGlnHisGlyArgLeuArgThrGlyPro 671
1907 GATATTCATAGCTGCCAATTCCTTAATGAAGCTTCAAACTGGATCC 1956
672 PheCysLeuProValSerValAspGlnProProSerTyrSerValle 688
1957 TACTGTCTCCACAGTGTGCTTGAAGAAATTCACCACTACCTCATGCA 2006
688 uThrProAsp.....ValAlaLeuProGlyMetArgTyrProValAspG 702
2007 TTCTGCGCAAGAAAGTCCCATTTACAGAAATCTCCCATTAAGTGGCTGAAG 2056
702 lYnHisLysGlyValPheSerValGlnLeuThrAlaValSerSerValHis 718

```

```

2057 GACATPAGGGAGATTATTAATTGAATGCAAGCTTTCTTCTGTACAC 2106
719 ProGlnAspProTyrLeuAspLysPhePheThrLeuValHisValLeuG 735
2107 ACCCAGCAACACCACCTGGAGAGATTTTCAACCTCTCCCTCCACTGGA 2156
735 u...GluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGln 751
2157 GAGCCAGGAGACCTTCCCATCCGCGCTGTGATCGAAGAAATCGAGGAGA 2206
751 lYAsnValGlnGlnLeuLeuArgAlaSerLeuAlaAlaLeuArgLeuAla 767
2207 TGGCCTGGAGCATGAGTGAAGCTCAGCATATCATCTGCTCAATCCTCC 2256
768 SerProGlnProLeuValAlaPheSerHisHisValLeuAspLysLeu 784
2257 GCGCTGGAGCGCGCTGTGCTCTCTCGACACCTGGTGGCGAAGCTCTT 2306
784 lArgLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeu 801
2307 CCAGCTGTCCGTGACAGCCCATGTGTCATGCTGGCCGACAGCAACTTCT 2356
801 lYArgGlyAlaPheGlnAlaMetAlaHisValValSerLeuValHisArg 817
2357 CCAGATTGCTTGCAGTCCGTGCTGCGCATCGCAACAGTCTGACACAC 2406
818 SerLeuGluAlaAlaGlnAspAlaArgGlyHisLysProGlnLeuAlaAl 834
2407 AGCAAGACACTGACAGAGACAGCATGAGAGAACTCCCTGCTGCTTC 2456
834 aTyrValHisTyrAlaPheArgLeuProGlyThrGlnProSerLeuPro. 850
2457 CTACGTGACACTGCTTCCGCTGCTCCAGAGTGCAGAAAGGATGTGCCCA 2506
851 ..AspGlyAlaProProValThrValGlnAlaAla.....ThrLeu 863
2507 ACTCAGGCGCTCCACTGCTCCTAGACCTCGGAGCTACCACACGATAT 2556
864 AlaArgLysSerGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSe 880
2557 GCGCCGACATCAGCTGCTGTGTGAGTTCAAAAGTCGTCGAGCGCGGCT 2606
880 rIleSerSerSerAsnProAspLeuAlaValAlaProGlySerValAsp 897
2607 GATGAGCAGCATTAACCACTCGCGGGGACACACTCCGACAGACAGC 2656
897 spGluValSerArgIleLeuAlaSerLysLeu..... 907
2657 AGAAGTGAAGAACATCATGTCTTCAAGATCGCGATCGCAACTGCAGC 2706
907 ..... 907
2707 CGAATGCTTACTATTGCTGTGGCAGTAGTGATGCTCCAAATTGCACGTC 2756
908 .....LeuHisGlnGluLeuAlaLeuG 915
2757 AGCCCAAGAGCCAGCCAGCAAAAAGCATTTCCATGAGAGAGCTTGCCCTTC 2806
915 lntTrpValValSerSerSerAlaValArgGluAlaIleLeuGlnHisAla 931
2807 AGATGTGTGTACGACCGGA..... 2826
932 TrpPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeu 948
2827 .....ATGTTGAAGAGCATGGCCGACGACGATACAAATAA 2858
948 uGlyAlaArgLeuAspThrProArgLysLeuArgPheProGlyArgPheL 965
2859 CATGGAACAAGCGGACAGCTTTTTCGAGAGACATCGTTTCTGACCGTTTCA 2908
965 euAspAspIleThrAlaLeuValGlySerValGlyLeuGlnVal..... 979

```

2909 TGGATGACATACACTATTGTTAATGTGGTCACTCGAAGAAATTCAGCC 2958
980 ... ILeThrArgValHisLysAspValGluLeuAlaGluHisLeuAsnAl 995
2999 CTTTAGTAAGAACACAGAGAAATGATGACGGGAAAGATGAAACAT 3008
995 aSerLeuAlaPhePheLeuSerAspLeuSerLeuValAspArgGlyP 1012
3009 CAGCTGGCTTCTCTTGTATGACCTTCTCCCTCATGGATGGGGCT 3058
1012 heValPheSerLeuValArgAlaHisTyrLysGlnValAlaThrArgLeu 1028
3059 TTGTTTAACCTCATCAGCATATTATTCAGCCAGCTGACGCCAACCTC 3108
1029 GlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMetGluPheTh 1045
3109 AGTAACCTTCA.....ACGCTATTCCATGAGGCTAGAGTTCCT 3149
1045 rArgIleLeuGlySerHisGluHisTyrValThrLeuAsnLeuProGlySc 1062
3150 GAGATCTCTGTAGCCATGAGCATTTACCTCAATCTGAACTTTTATTA 3199
1062 ySProLeu...SerProAlaSerProSerProSerValSerSerThr 1077
3200 TGAATGCTGATCTGCTCCACATCTCTCTGCTCTCCATATCT..... 3243
1078 ThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLysValTh 1094
3244 ...TCCAGAACTCAAGCTCCGCTCCAGCTTCACAGAACAGAAAGATCGC 3290
1094 rSerMetPheGluLeuSerGlyProPheArgGlnGlnHisPheLeuAlaG 1111
3291 CAGATCTGTGATCTGACTTCGAGTCCGAGCCAGCACAGCACTTCTCAGCG 3340
1111 LysLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaGluGly 1127
3341 GGCCTCTTCACAGAACTGGCTGCTCCCTGAGTGGCGAAGGGAGGA 3390
1128 AlaPheLeuLeuHisLysLysAlaIleSerAlaValHisSerLeuLeuCy 1144
3391 ATCAGAAAGTACAAAGAAAGCTGTGCAATTCACAGCTCGCTAG 3440
1144 sGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAlaArgV 1161
3441 TTCACGACCTGGAGCCACGCTGTGTCAAAACAGAGGTGAAGGTCAAAA 3490
1161 alaAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuPro 1177
3491 TCGCGCGCTTACCTACCTTACCTTACCTTACCTTACCTTACCTTACCT 3540
1178 ArgLeuHisAspPheAlaGluGlyProGlyGlnArgSerArgLeuAlaSe 1194
3541 CAGCTCTGTGCTTACAGTTGAGATTCGAGATACCGACACAGCTGG 3590
1194 rMetLeuAspSerAspThrGluGlyGluGlyAspIleAlaGlyThrIleA 1211
3591 C.....TCGGATGAAGAAACAAAGAA.....GCCGGTGCACATTA 3625
1211 snProSerValAlaMetAlaIleAlaGlyLysProLeuAlaProGlySer 1227
3626 ACCGAGATGTGGCTCTGGCCATAGCAGGAAATATTTCATTTGAAGAAC 3675
1228 ArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGlyCysAl 1244
3676 AGTGAGATAGTGTCTCTCTCTGCC.....TATAGACATGACAAACAT 3719
1244 aLeuSerAlaGluSerSerArgThrLeuLeuAlaCysValLeuTyrValL 1261
3720 GCTGAAGCGGACACTACTCGCAACCTCATGATCTGCTCTGTGATCA 3769
1261 euLysAsnThrGluProAlaLeuLeuGlnArgThrAlaThrAspLeuThr 1277
3770 TGAATAATGCTGATCAGACCTCTATTAGGAATGGATTTGCTGACCTGCC 3819

1278 LeuProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuGlyLeuAlaAl 1294
3820 TCAAGCCACTCATCAGAGATTTTAGATCTACTTTTCATCTGTGTATG 3869
1294 aPheGluTyrLysGlyLysLysAlaPheGluArgIleAsnSerLeuThrP 1311
3870 TTTTGGATTAAGGGGAAACACAGATCTGACAAAGTCAAGTACCAAGTCC 3919
1311 heLysLysSerLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGly 1327
3920 TCGAGAGTCAAGGGATGTCAAGGCCCGCTGGAAGAGCTTTGCTGCGT 3969
1328 ThrIleGlyAlaArgGlnGluMetValArgArg.....SerArgG1 1341
3970 GGGGAAGGGGCGCAGAGGGAGATGATGGCGCCGGGCTCCAGGGAAGA 4019
1341 uArgSerProPheGlyAsnProGluAsnValArgThrArgLysSerValT 1358
4020 CCGATTTCCA...GGCCTTAATGAAATTTGAGATGGAAGAAAGACAGA 4066
1358 hrHisTyrLysGlnThrSerAspArgValAspLysThrLysAspGluMet 1374
4067 CACATTGGCGGCGACGCTAATGGAAGCTAGATTAACAAAGCCGAGTTA 4116
1375 GluHisGluAlaLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuVa 1391
4117 GATCAAGAAAGCTTATGATGAGTGCATCTGGCTACAGAAACATTTAAT 4166
1391 lValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGluA 1408
4167 CATTCTGGATATGACAGAAACATTTATCCAGGCGAGCTCGGCTTGACT 4216
1408 lArgGluSerValLeuGluAlaValAlaLeuLysValValLeuTyrSerLeu 1424
4217 GTAAGACAGCCTGCTGGAGGATGTTCTGAGGCTGCTGTAATCTCTG 4266
1425 GlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnAr 1441
4267 AACTGTGACAGAGTACACCTACTCTGACTCACTCTTGCACACCTCGC 4316
1441 gAlaLeuValSerLysPheProGluLeuLeuPheGluGluAspThrGluL 1458
4317 TCTCTCATGCGCAAGTTTGGAGACTTACTTCTTCABAAGAGGTTGGAAC 4366
1458 euCysAlaAspLeuGlySerLeuArgLeuLeuArgHisCysGlySerArgIle 1474
4367 AGTGTTTGACCTATGTACACAAAGTCTGACACCATGACAGCAGCATG 4416
1475 SerThrIleArgThrHisAlaSerAlaSerLeuTyrLeuLeuMetArgG1 1491
4417 GATGTACCCGAGGACCAAGCCTGTGCCACCTTACCTCTCATGAGGTT 4466
1491 nasnPheGluIleGlyHisAsnPheAlaArgValLysMetGlnValThrM 1508
4467 CAGTTTGGAGCCACAGTAATTTTGGCAAGTAAGATGCAAGTAACCA 4516
1508 eSerLeuSerSerLeuValGlyThrThrGlnAsnPheSerGluGlnHis 1524
4517 TGTCCCTGGCATCTTGTGGGGAAGACAGCAGACTTAAATGAAGACAGC 4566
1525 LeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluGluAspMetG1 1541
4567 CTGAGAGATCTCTTCCACCCAGCTGAGAACTTCTGTATATC 4666
1541 yLeuArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnL 1558
4617 CATGAGATGACTCTTTCCACCCAGCTGAGAACTTCTGTATATC 4666
1558 euHisMetIleLeuThrAspThrValLysMetLysGlnHisGlnGluAsp 1574
4667 TGAATAGCATCTTATATGACACAGTGAATAATGAGGAATTTTCAGGAGAT 4716

```

1575 ProGluMetLeuIleAspLeuMetYrArgIleAlaArgGlyTyrGlnG 1591
|||||.....|
4717 CCTGAGATGCTTATGAGATCTCATGTACAGAAATGCCAAGAGTACCAGGC 4766
ySerProAspLeuAvrLeuThrTrpLeuGlnAsnMetAlaGlyYrHisA 1608
:|||||.....|
4767 AATCTCTGATCTGCGCTGACCTGCTCCAGAACTGGCAGAGAAACACA 4816
1608 IagLLeuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaA 1624
:|||||.....|
4817 CCAAGAAAGAAAGTGTACACAGGAGGCTGCATGTGCTGTGCACGCCCT 4866
1625 AlaleuValAlaGluTyrLeuAlaLeuLeuGlnAspGlnArgHisLeuP 1641
|||||.....|
4867 GCGTTAGTGGCTGAGTATCTGACGACTGTGAGAGACCCACACTTACTGCC 4916
1641 oValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuGlnGus 1658
|||||.....|
4917 CGTGGGCAGTGCTGCTCCAGAAATATTCTTCCAAATGCTGCTGGAGAGT 4966
1658 eValIleSerAspAspIleLeuSerProAspGlnGlnGlyPheCysSer 1674
|||||.....|
4967 CTGTGCTCTGAGGACACCTGTCTACTGACGAGAGATGGGGTGTGCACA 5016
1675 GlyYrHisPheThrGluLeuGlyLeuValGlyLeuLeuGlnGlnAlaA 1691
|||||.....|
5017 GGCACGTACTTACACGAGAGTGGCTGTAGCCCTCGGAGACAGGCCGC 5066
1691 aGlyYrPheThrMetGlyGlyLeuTyrGluAlaValAsnGluValTyrL 1708
|||||.....|
5067 GGAGCTCTTGACGACGGAGGCTTATATAGACAGTTAATGAGGCTTACA 5116
1708 yAsnLeuIleProIleLeuGlnAlaHisArgAspTyrLysLysLeuAla 1724
|||||.....|
5117 AGCTGTCATCCCATCTCTAGACGCGCATCGAGAAATTCGGAGAGGTGACA 5166
1725 AlavaLHisGlyLysLeuGlnGluAlaPheThrYrHisIleMetHisGln 1741
|||||.....|
5167 CTCACTCACAGCAAGCTCGACAGAGACCTTCGACAGCATCTTAACAAGGA 5216
1741 rSerGlyTrpGluTrpValPheGlyThrTyrPheArgValGlyPheYrG 1758
|||||.....|
5217 TCAT.....AAGAGAAATGTTTGAACCTACTCTCGAGTGTGTTCTTGG 5260
1758 lYAlaHisPheGlyAspLeuAspGlnGlnGluPheValTyrLysGluPro 1774
|||||.....|
5261 GATCCAAATTTGGGGATTTGGATGATGACAGAGGTTTGTCTACAAAGAGCCT 5310
1775 SerIleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheYrTh 1791
|||||.....|
5311 GCAATTTACCAAGCTTCTGAGATCTCACATAGACTAGAGGATTTTATGG 5360
1791 rGluArgPheGlyAspValValGluIleIleLysAspSerYrProV 1808
|||||.....|
5361 TCATATGTTTGTGTCAGAAATTTGhGAAAGTATTAAAGACTCCACTCTCG 5410
1808 aLAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrTyr 1824
|||||.....|
5411 TGGACAAAAACCAAGTGGATCTTAACAAGGCTTCATACAGATACACTTTT 5460
1825 ValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTyrPh 1841
|||||.....|
5461 GTGAGGCCCTTACTTGTAGATATGAGATGAAGACAGGAGGTCACATACTT 5510
1841 eAspArgAsnTyrCylLeuArgThrPheLeuPheCysThrProPheThrP 1858
|||||.....|
5511 TGAGAGAAATTTCAACCTCGGAGGTTCTATGACACACCCCTTGTCACCC 5560
1858 roAspGlyArgAlaHisGlyGluLeuProGlnGlnHisLysArgLysThr 1874
|||||.....|
5561 TGGAGGGGCGGCGCTCGGAGAGAGCTGCATGAGCAGTACAGAAAGAACACA 5610
1875 LeuLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArgVa 1891

```

```

5611 GTCTTGACCACTATGACAGGCTTCCCTTACATCAAGACCGATCAGCGT 5660
1891 lCySHisArgGlnGluThrValLeuThrProValGluValAlaIleGluA 1908
|||||.....|
5661 CATCAACAAGAGAGAGTTGTTTGGACCCGATTTGAAGTTGCCATTTGAAG 5710
1908 sPheGlnLysLysThrArgGluLeuAlaPheAlaThrGlnGlnAspPro 1924
|||||.....|
5711 ACATGAAGAAGAAGACCCCTGACATTGACAGTTCCTCAATTAACCGAGACCG 5760
1925 ProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyProh 1941
|||||.....|
5761 CTGATGCAAGAGATGCTTCAATGTGTGTGCAAGGCTGTGTGGAGCTAC 5810
1941 rValAsnGlnGlyProLeuGlnValAlaGlnValPheLeuAlaGluIleP 1958
|||||.....|
5811 TGTAATACAGAGGACCACTGAGATAGCCCAAGTGTGTTGGCTGAATATC 5860
1958 roGluAspProLysLeuPheArgHisHisAsnLysLeuArgLeuCysPhe 1974
|||||.....|
5861 CTGCTGATCCAAAACCTTATCGACATCACAAACAGTTGAGTTATCTTTT 5910
1975 lYsAspPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLe 1991
|||||.....|
5911 AAGGAATTCATCATGAGATGTGTGAAGCTGTAGAGAAACCAAGGCTCT 5960
1991 uIleGlyProAspGlnLysGluTyrHisArgGlnLeuGlnArgAsnTyrC 2008
|||||.....|
5961 CATCACGGCAGACACAGAGGAAATATCAGCAGAACTCAAAAAGAACTATA 6010
2008 yAsArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgLeuProGln 2024
|||||.....|
6011 ACAAGCTAAAGAGAAACCTCAGGCCATGATCGAGGGAATAATTCAGAA 6060
2025 LeuMetAlaPro.....ThrProGlyLeuArgAsnSerLeuAs 2038
|||||.....|
6061 CTGTACAGACCAATATTTCAGAGTTGAGAGTCAAAAGAGGACTCTTCCA 6110
2038 nArgAlaSerPheArgLysAlaAsp 2046
|||||.....|
6111 CAGATCTAGTTTCAGGAAATGTGAA 6135
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:ABLI17736
seq_documentation_block:
ID ABLI17736 standard; DNA; 9389 BP.
XX
XX ABLI17736;
AC
XX
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster genomic polynucleotide seq ID NO 4681.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
XX pharmaceutical; gene; ds.
XX
XX Drosophila melanogaster.
OS
XX
XX W0200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW.
XX
XX WPI; 2001-656860/75.
DR

```



```
6796 GGCAGCAGCTTTGGCACAAGCCCAAGCTCCAATAGCTTAGATCGAAAATC 6747
424 aTrpTh.....AsparGArGArGArglyProGlnasp..... 435
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6746 TTCACAAGTAGCTTTGATCACTAAAGCGAAAAGGTAAGATATGAGCG 6697
436 .....ArgAla 437
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6696 GAAACCTTACAGACGTGCTCGCTTGACGGAAGAAAAGCAACCGATCC 6647
438 SerSerGlyaspSer...AlaCysSerPheSerGlyPheArgProAlaTh 453
      ||| ||||| ||| :||| :||| :||| :||| :||| :||| :|||
6646 TGTGCCCCAATGATTTTGCACAACGTTGAGAAAACCTTCTGCTATCAC 6597
453 rLeuThValThrAsnPhePheLysGlnAlaGluArgLeuSeraspG 470
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6596 TATTAACAGTACCAAGCTTTTAAACAAGAGCAGATAAATGAAGACG 6547
470 LuAsrLeuPheLysPheLeuAlaAspMetArgArgProSerSerLeu 486
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6546 AGGATTTGTCAAAATTTTACCAAGAGTTAAACGACGAGCTCAGTTATG 6497
487 ArgArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAl 503
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6496 AAGAAAGTACAATCCATACCTGCTTATTAATTAATAATATACACTTG 6447
503 aProGluAsnProHisPheCysLeuSerProGluLeuLeuHisIleLysP 520
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6446 CGTCGAGGAGCCAGAAATGATTAACGCGAGTTAGCAGCTATGAAC 6397
520 roLyArgProAspArgGlyArgProThrLysGlnIleLeuGluPhePro 536
      || :||| :||| :||| :||| :||| :||| :||| :||| :|||
6396 CTCAAACTGCAGATAAATGCGCCAGTTAAGGAATCTTAAGATTTCCG 6347
537 AlaArgValAlaTyralaProHisThrSerTyArgAsnLeuLeuTyVa 553
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6346 CAATCGCGCATTTTCATCCATCCATTAAGCTATCGCATCTCTATTGT 6297
553 lTyTrpHisSerLeuAsnPheSerSerArgGlnGlySerValArgAsnL 570
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6296 GTCCACCAAGAGCTAAATTTTCTCGCGTGCAGGTTCTGCTCGAACA 6247
570 euAlaValArgValGlnTyMetThrGlyLysPheProSerGlnAlaLeu 586
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6246 TTGCCGCTGCGGTCAACTAATGGCTGGAGAAAGCCAAAGATGCGGTC 6197
587 ProValIlePheGlyLysSerSerCysSerGluPheThrArgGlnAlaPh 603
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6196 AATGCCATTCACGCGCAGTGTCTGTGCCAATTTCTTACTAGACATTT 6147
603 eThrProValValTyrlHisAsnLysSerProGluPheTyGluGluPheL 620
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
6146 TACAGCTGCAATTAATCAATAAATGCCATCATCTTACGATGAATAA 6097
620 ySLeuHisLeuProAlaCysValThrGluAsnHisIleLeuPheThr 636
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6096 AAATTGGCGTGCCTGCATCAATAAGCAGCATCACCATTTATTTGTAC 6047
637 PheTyrlHisValSerCysGlnProArgPro.....GlyThrAlaLe 650
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
6046 ATTATCCATGTTTCATGTCAAAAGAAACACAGAGACTTGCAGCGCTC 5997
650 uGluThrProValGlyPheThrTrlPheProLeuLeuGlnHisGlyArgL 667
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5996 AGGACACCTCATCGCTACACATGTTGCGCTTACTGGAAGATGGAAC 5947
667 euArgThrGlyProPheCysLeuProValSerValAspGlnProProP 683
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5946 TTAAGTTGGGGAATTTAATCTTCTGTTATGTTAGATACCGCGCGAA 5897
684 SerTyrlSerValLeuThrProAspValAlaLeuProGlyMetArgTrpVa 700
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5896 AATTAGCATTTTATACACCGCAATGTTCACTGCTGGAATTAATAGGCT 5847
```

```
700 LuAspGlyHisLysGlyValaPheSerValGluLeuThrAlaValSerSer 717
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5846 GGACATACATCAGACGCGGTGTTTCTATTAAATAGTAGAAGCAGTACCG 5797
717 aHisPProGlnAspProTyrlLeuAspLysPhePheThrLeuValHisVal 733
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5796 TTCAATACCTTGGATCTTCTCTCGATGATGATCTTCTTAATTTGCGAAT 5747
734 LeuGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerG 750
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5746 CTAGACACTCGAAATTAATCT.....TCCCATATTGGCGCA 5712
750 uGlyAsnValGluGlnGluLeuArgAlaSerLeuAlaAlaLeuArgLeu 767
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5711 GAATATATAGAAACAGAAATTAAGAAATGCTGCTGATATCGCAATATG 5662
767 LuSerProGluProLeuValAlaPheSerHisValLeuAspLysLeu 783
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5661 CAATCTGTAGCCATTATGAGGCAATCTACCATTAAGTCTTGATTAAGCT 5612
784 ValArgLeuValIleArgProPheIleSerGlyGlnIleValAsnLe 800
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5611 ATCGAATTTGTTAGTTGTCCACCACAAAGTTGGTGCACAAGCAATGCT 5562
800 uGlyArgGlyAlaPheGluAlaMetAlaHisValaValSerLeuValHis 817
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5561 GGGATCTACGGTTTTCGAAGTCTTTGTTGGTATCATCTTTGTTATCA 5512
817 rGSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAl 833
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5511 TACTTAACGACATCAATATAGCAATACGGTGGCAAAAGTTGGCTATCA 5462
834 AlaTyrlValHisTyralaPheArgLeuProGlyThrGluProSerLeuPr 850
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5461 ACATATGTCACCTTCAATGTAATTCGCCA..... 5430
850 oAspGlyAlaProProValThrValGlnAlaAlaThrLeuAlaArg... 865
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5429 .....CCCTTCAAGCAACACGAGCTACATGACGTGCGC 5393
865 ..... 865
5392 AGCACACAGAAAGCTTGCATTAAGCGAATCGTACACCATATATGATTA 5343
866 .....GlySerGlyArgProAlaSerLeuTy 874
5342 TGTTTTAGCAATGGCGCGAAGCTTGATGCGAAAG...TAAATACATA 5296
874 rLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAlaVala 891
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5295 TTTTGTAAAGGCTTAAATATATACCTTTTCTTTCACAGAGCTGCATA 5247
891 LuProGlySerValaAspAspGluValSerArgIleLeuAlaSerLysLeu 907
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5246 .....GATCTCTTCACTGATGCTGCACGAGATGTCCAAGTGGCGCTG 5202
908 LeuHisGluGluLeuAlaLeuGlnTrpValaValaSerSerAlaValaAr 924
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5201 CTTCAGAAAGATTTAGCCTTCATTTGGGTTCGCAAGTGGAAAGGCGAGC 5152
924 gGluAlaIleLeuGlnHisAlaTrpPhePheGlnLeuMetValLys 941
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5151 TCACCTTAGCATGTCAAATTCGTGCTTTTATTTAGCTTATTTGCAAT 5102
941 ePheAlaLeuHisIleLeuLeuGlyGlnArgLeuAspThrProArgLys 957
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5101 CTATGATTGAACATTTCATTGTTCAAAATCTTAAATGGTCCCGCGAA 5052
958 LeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValaGly 974
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
5051 CATGATTTCCACATCAGTATTAAATGATGATCATCCACCCTGGTGCAC 5002
```


3271 GCGGTGGCTTAACCGTCTGTTATGCTAATCCGACTCCGACTG 3222
1543 ArgAspSerThrPheAlaGluGlnValGlnAspLeuMetPheAsnLeuH1 1559
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
3221 CAGGAACTCTTCTTCCAGCAAGTACAAGATTCTTTTAACTGCA 3172
1559 SMetIleLeuThrAspPheValLysMetLysGlnHISGlnGluAspProG 1576
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
3171 TATGATCTCTGATGATCTGTTAAATGAAGGATGTCAGGAAGACCCAG 3122
1576 IuMetIleAspLeuMetTyrArgIleAlaArgIleTyrGlnGlySer 1592
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
3121 AATGTTGCTGACCTCATGAAATCGTATGCCCAGGATACCAAAATAAC 3072
1593 ProAspLeuArgLeuThrTyrPleuGlnAsnMetAlaGlyLysHisAlaG1 1609
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
3071 CCTGATCTGACTGACTGTTGGAATAATGCTAATAAACAACCCGCA 3022
1609 uLeuGlyAsnHisAlaGlnAlaGlnCysMetValHisAlaAlaAlaL 1626
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
3021 GCGAGCAAAATCACGGAAGACGCCATGTGTATGTACATGCTGCTTCT 2972
1626 euValAlaGluTyrLeuAlaLeuLeuGlnAspGlnArgHisLeuProVal 1642
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2971 TAGTTCTGAATATCTTAGCATGTGTGAGTCACAAAACATTTGCCCTGT 2922
1643 GlyCysValSerPheGlnAsnIleSerAsnValLeuGluGluSerAl 1659
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2921 GGAGCTGTAAATTTCACGAAATTTCTCCCAACCACTAATGAGCGGC 2872
1659 aIleSerAspIleLeuSerProAspGluGluGlyPheCysSerGlyL 1676
|:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2871 CGATGCGATGATGCTGTAAGTCCCGCAAGATGTATCTGCTTAGGAA 2822
1676 yHisPheThrGluLeuGlyLeuValGlyLeuLeuGlnGlnAlaGlyL 1692
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2821 ATCATTTCACTGAACCTGGGTTGAAGCCTTGCTGGAAGAGCCTCCAT 2772
1693 TyrPheThrMetGlyLeuTyrGlnAlaValAsnGluValTyrLysAs 1709
|||:::|||||:::|||||:::|||||:::|||||:::|||||
2771 TCTTTCAGATTGCTGCATGATGAAGCAATGACGAAGTGTACAAAT 2722
1709 nLeuIleProIleLeuGlnAlaHisArgAspTyrLysLysLeuAlaVal 1726
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2721 TCATAATCCCATATGCGAGGCTAACAGAGATTTCAAAAGCTAAGCAAG 2672
1726 aHisGlyLysLeuGlnGlnAlaPheThr..... 1735
2671 TTCATGGCAAAATGCAAGAGGCAATTAAATCCCACTACAGATA 2622
1736LysIle.MetHisGlnSerSerg 1743
2621 ACAATATGTGTAAATTTTACCAACGAAACAAATATATATTTATAAAC 2572
1743 LyrProGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAla 1759
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2571 AGGCTAAGAGAGTTTGGAACTACTTTCGTGTTGCTTCTTAGTGGCGGA 2522
1760 HisPheGlyAspLeuAspGluGlnGluPheValTyrLysGluProSerL1 1776
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2521 AAATTTGGGACTT3GATCAGCAGAAATTCATTATTAAGAGCAACAT 2472
1776 eThrLysLeuAlaGluIleSerHisArgLeuGlu..... 1787
2471 GACGAAGTTGCCGAAATATTTAGTGGCTCAGGTATATATTTGCAAAAT 2422
1788Glu.PheTyr 1790
2421 GGAATAAATAGAACTAATCAATTTTTCACAAATACGTTAGACTTTTA 2372
1790 rThrGluArgPheGlyAspAspValAlaGluIleLeuLysAspSerTyrP 1807
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2371 CACTGAACGATTCGACGAGACTCTGTCATATCATTAAGATTCACATA 2322

1807 roValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThr 1823
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2321 CCGTGATATTAATAGCTTGGATCCCGATTAAGGCTTACATTCAAATTA 2272
1824 TyrValGluProTyrPheAspThrTyrGluLeuLysAspArgValThrTy 1840
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2271 TATGTTGAACCTTACTTTGAAACATATGAAATGCGTCATCGTGAACATA 2222
1840 rPheAspArgAsnTyrGlyLeu..... 1847
2221 CTTTACGCGGAATTCATATATAGATGATATGAAATTAACAGATAATT 2172
1848ArgThr 1849
2171 AAATGAAATTTAAATATATATTTAATTTGTTTGTATGCAAGAC 2122
1850 PheLeuPheCys..ThrProPheThrProAspGlyArgAlaHisGlyGlu 1866
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2121 TTTTATATATGCGACGCTTTTACTAAAAAATGAAGGACACATGCGAAC 2072
1866 euProGluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1882
|||:::|||||:::|||||:::|||||:::|||||:::|||||
2071 TAAATGACAGAGTAAACGAAACGATTTTGACCGCGGCAATCACTT 2022
1883 ProTyrIleLysThrArgIleArgValCysHisArgGluGluThrValLe 1899
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
2021 CCATCGTAAAAACCGCATATGTTATGACCGACAGCAAAATAGTCT 1972
1899 uThrProValGluValAlaIleGlu..... 1907
|:::|||||:::|||||:::|||||:::|||||:::|||||
1971 GCAACCGATTGACGTGCAATTAAGATTAAGTATGTTGTAATAATCCAAA 1922
1908AspMetGlnLysLysThrArg 1914
1921 AGTTATATAAATTAACGTATGCGCTAGATATTCAAAAAACAATG 1872
1915 GluLeuAlaPheAlaThrGluGlnAspProAspAlaLysMetLeuG1 1931
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1871 GAATGGCAGCTGCCCAATCAAGAGCCAGACGCCCAAAATATGCA 1822
1931 nMetValLeuGlnGlnLysSerValGlyProThrValAsnGlnProLeuG 1948
|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1821 AATGCTACTGCAAGGATGATTTAGCAACCTGTTAACCAGACCGATGG 1772
1948 IuValAlaGlnValPheLeuAlaGluIleProGluAspProLysLeu... 1963
|:::|||||:::|||||:::|||||:::|||||:::|||||
1771 AAATGCGCAGTGTCTTCCATTAATCCAGACGAAACATGACCG 1722
1964 PheArgHisHisAsnLysLeuArgLeuGluCysPheLysAspPheCysLysL 1980
:::|||||:::|||||:::|||||:::|||||:::|||||
1721 ACAAAACACCAAAACAAACTTCGGTTGCTTGCAGCGATTTCAAGCG 1672
1980 sCysGluSproAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnL 1997
|||:::|||||:::|||||:::|||||:::|||||:::|||||
1671 TTGCTCTATGCTTTGAAAAGAAATCGCAATCTATATCTTCAATCAADA 1622
1997 ySGluTyrHisArgGluLeuGluArgAsnTyrCysArgLeuArgGluAla 2013
|||||:::|||||:::|||||:::|||||:::|||||
1621 AAGATTACCAACGAGAACTGAGACGTAAACAACGATCGGTTCAATTGAACA 1572
2014 LeuGlnProLeuLeuThr 2019
|||:::|||||:::|||||:::|||||:::|||||
1571 TTGACTCCCTTATTAATCT 1554
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAH41953
seq_documentation_block:
ID AAH41953 standard; DNA: 4200 BP.
XX AAH41953;
AC
XX
DT 30-AUG-2001 (first entry)

[illegible]

CC related sequences can be used in preventing or treating a CIASP-3 mediated disease, preferably an autoimmune disease by inhibiting an immune response. The autoimmune disease is caused or exacerbated by increased activity of T helper 1 lymphocytes (TH1). AAH41893 to CC and AA899491 to AA899507 represent sequences which are used in the exemplification of the present invention. CIASP-3 is localised in the CC chromosome location 1p31.1.

XX Sequence 4143 BP; 1251 A; 829 C; 890 G; 1173 T; 0 other;

alignment_scores:

Quality: 4273.00 Length: 1247
Ratio: 4.039 Gaps: 11
Percent Similarity: 84.844 Percent Identity: 67.121

alignment_block:

US-09-736-968a-2 x AAH41911 ..

Align seg 1/1 to: AAH41911 from: 1 to: 4143

```
877 ArgSerLysSerLieserSerSerAaspProaspLeuAlaValAlaProG1 893
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
10 CGTTCGGAAGCCTTAGTATATACATCCAGATATATCTGGGACTCCAC 59
893 ySerValAspAspGluValSerArgLleLeuAlaSer..... 905
||| |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
60 GTCACAGATGATGAGTTCGATCATCATCGGAGTAAAGGTTGATGATC 109
905 ..... 905
110 GCTCCATTCTCTGGTTCACACTGTGTCCAAAGCTGCCCATGGGA 159
905 ..... 905
160 TCCACCCAGTCCAGTGCAGATCAACACAGCTATGATGAGATG 209
905 ..... 905
210 TAATCGTAGTCTTCGCACACAGACGTCACGTTTCTACAAACATTA 259
906 ..... 905
260 CGGAGCGCTTACCACTAAAGCTTTTTCACGAGAGAGCTGGCTTGCAG 309
916 TrpValAlaSerSerSerAlaValArgLleAlaLeuGlnHisAlaTr 932
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
310 TGGTGTGTTGAGTGAGCGACGTTTCGGGAATCAGCTTGGCAACACCTG 359
932 pPhePhePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuG 949
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
360 GTTCCTTTTGAATTAAGGTAAAGAGCATGTCACATTTATCTTAA 409
949 LysLlnArgLeuAspTrnProArgLysLeuArgPheProGlyArgPheLeu 965
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
410 ATGATTAACCTTGAGGCTCCAGGAAAAGTCGTTTCCACAGACCTTTCATG 459
966 AspAspLleThrAlaLeuValGlySerValGlyLeuGlnValLleThr 982
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
460 GATGACATTCAGCTCTGTTCACAGCATGCTAGATATAGTTTCAAG 509
982 gValHisLysAspValGluLeuAlaGlnHisLeuAsnAlaSerLeuAlaP 999
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
510 ATTTCAGAAAGACACAGAAATGGTTAGAGACATCAATACAGCCTTGCA 559
999 hePheLeuSerAspLeuSerLeuValAspArgGlyPheValPheSer 1015
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
560 TCTTTCATATGATCTGTCTGTATGACAGAGATTTGTTTATAG 609
1016 LeuValArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSerSerP 1032
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
610 CTATATAAGTCCTGATATAACAGGTGCTTCAAGCTTACTCATTAAC 659
```

```
1032 CAsnProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgLleLeu 1049
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
660 GAATCCAGCTTCTGCTGTCCTTGAGGCTGATTTCTPAGAAATCATCT 709
1049 ySerHisGlnHisTyrValThrLeuAsnLeuProCysCysProLeuSer 1065
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
710 GCAGTCATGACGACATATGTATTAATTAACCTGACGCTTACTTACT 759
1066 ProProAlaSerProSerProSerValSerSerThrTrnSerGlnSer 1082
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 CCACCTGCATCTCCATCACCCTTCTGTTCTCTGCACATCTCAGAGTTC 809
1082 rThrPheSerSerGlnAlaProAspProLysValThrSerMetHeGlu 1099
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
810 TGGATTTCTACGAATGTACAAAGAACCAAAAGATTCGATTTGTTGAT 859
1099 eUeSerGlyProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThr 1115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
860 TATCCGTCCTTTCGCGCACAGCATTTATTTGGCAGACTGTGTATACA 909
1116 GluLeuAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuHis 1132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
910 GAGCTGGCTGTCATTTTAGACCCGTGATGCTGAGAGACTGTTGGATTGCA 959
1132 sLysLysAlaLieserAlaValHisSerLeuLeuGlyHisAspThr 1149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
960 TAAAGAGTCATCAATATGATGTACACAAATTTACTCTCCGTCACACATCAG 1009
1149 sPProArgTyrAlaGluAlaThrValLysAlaArgValAlaGluLeuTyr 1165
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1010 ACCCGGCTGACTGCTGACCTCAGATTAAGGCTCGAGTGGCATGTGTAT 1059
1166 LeuProLeuLeuSerLleAlaArgAspThrLeuProArgLeuHisAspPh 1182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1060 CTACCTGATGTGATATATCATGGAACCTGATCCTGATGATTT 1109
1182 eAlaGluGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSera 1199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1110 TACAGAACTCACATCAACAGGAGGAAACCAATTTGTATGCCACTGATG 1159
1199 sPThrGluGlyGluLysAspLleAlaGlyThrLleAsnProSerValAla 1215
|| ||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1160 ATTATGAAGTAGAGC.....GGAAGTATGATAGCAACCGCTGCC 1203
1216 MetAlaLleAlaGlyProLeu.....AlaProGlySerAr 1228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1204 ATGGCAATCGCAGGAGACATGCTCCCTCACTAACAGAGCTGCAAGTTT 1253
1228 gAlaSerLieserGlnGlyProProThrAlaSerArgLleGlySAlaL 1245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1254 CCTCTCAGCTCA.....ACGAGTGCAGGCGACACACTACCT 1291
1245 eUeSerAlaGluSerSerArgThrLeuAlaCysValLeuTrpValLeu 1261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1292 TTTTCAGCAAGATCAAGTCGAAAGCTTTTGATCTGTCTACTTGGGTTCTC 1341
1262 LysAsnTrnGluProAlaLeuLeuGlnArgTrpAlaThrAspLeuThr 1278
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1342 AAAAATGCAGATGAAGACAGTCTACAGAAAGTGTATACAGATCTCAGT 1391
1278 uProGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuGlySAlaAlaP 1295
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1392 CTTCGAGCTAAACCGGCTATTAGATCTCTTATCTCTGTGTGCTTGCT 1441
1295 heGlyTyrLysGlyLysLysAlaPheGluArgGlyLeuAsnSerLeuThrPhe 1311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1442 TTGATATATAAAGGAAAAAGTGTGAAAGCAATGAATGATGACCTTT 1491
1312 LysLysSerLeuAspMetLysAlaArgLeuGlnGluAlaLleLeuGlyTh 1338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1492 AAGAAATCAAAAAGCATGAGCAAAAGCTTGAGAGACTATTTCTGGAG 1541
1338 rLleGlyAlaArgGlnGluMetValArgArgSerArg.....GluA 1342
```



```
1542 CATAGTCCAGGCAAGAAATGTAACGCCGAGACAGCTCCAGA 1591
      |||
1342 rgsrPro.....PheGlyasnProGluasnValArgTrrArg 1354
      |||||
1592 GAAGCCCATCTGGAAAGTCCCTTGGAAAGTCAGAAAATTTGAGGTGGAGG 1641
      |||||
1355 LysSerValThrHisTrpLysGlnThrSerAspArgValAspLysThrLy 1371
      |||||
1642 AAAGTATGACTCACTGCGCTCAAAACACAGAGAAGCTTGCAAAATCAAG 1691
      |||||
1371 sAspGluMetGlnHisGluAlaLeuValGluGlyAsnLeuAlaThrGlu 1388
      |||||
1692 AGCAGAGATTGAACACGACGATGATGATGAACCTGGCTACACAAG 1741
      |||||
1388 LasSerLeuValLeuAspThrLeuGluLeuIleValGlnThrValMet 1404
      |||||
1742 CAACCTAATCATTTAGATATCATAGATGTTGTTCCAGACCGCTTCT 1791
      |||||
1405 LeuSerGluAlaArgGluSerValLeuGlyAlaValLeuLysValVal 1421
      |||||
1792 GTACCGGAAATCCAAAGAGACATTCCTGGTGAGCTGCTAAAGAGTCTACT 1841
      |||||
1421 uTrpSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeu 1438
      |||||
1842 ACACGCGATGGCTGTAAACCAAGTGCAGTTATCTACAAACACTGTTTG 1891
      |||||
1438 LathrGlnArgAlaLeuValSerLysPheProGluLeuLeuPheGluGlu 1454
      |||||
1892 CTACACAGAGAGCCTTGCTTCAAAAGTTCTCTGAACCTTATTTGAAGAA 1941
      |||||
1455 AsphTrpGluLeuCysAlaAspLeuCysLeuArgLeuLeuArgHisCysG 1471
      |||||
1942 GAGACAGAGCGATGCTGATTTATGCTCAGCGCTCTCCGACACTGTAG 1991
      |||||
1471 YSerArgLLeSerThrIleArgThrHisAlaSerAlaSerLeuLysLeuL 1488
      |||||
1992 CAGTAGCATTCGGTACATACGGTACACCCAGATGCTCCCTTAACTTAC 2041
      |||||
1488 eumetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504
      |||||
2042 TAATGAGGCAAACTTGGAGATGGGAATACTTGGCAGGGTTAAATG 2091
      |||||
1505 GlnValThrMetSerLeuSerLeuValGlyThrThrGlnAsnPheSe 1521
      |||||
2092 CAGGTACCAATGTCATATCTCTCTGTTGGTGGCAGATCTCAGAAATTTAA 2141
      |||||
1521 rGluGlnHisLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGluG 1538
      |||||
2142 TGAAGAAATTTTAAAGCGTCTCTAAAGACTATATGTGACATATGCTGAAG 2191
      |||||
1538 LuAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLeu 1554
      |||||
2192 AAGATCTGGATTTGAGGGAACAACATTTCTCGATCGACGTCAGAGATCTG 2241
      |||||
1555 MetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGluH 1571
      |||||
2242 GTTTTCAATCTCCATATGATTTCTTGTGATACGTAAGTGAAGGAAACA 2291
      |||||
1571 sGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAlaArg 1588
      |||||
2292 CCAGGAGATCTGAAATGTTGATGTATGATGTACAGAAATGCCCCAAGG 2341
      |||||
1588 LyrTrpGlnGlySerProAspLeuArgLeuThrTrpLeuGlnAsnMetAla 1604
      |||||
2342 GTTACGAGACTCTCCAGATCTGGGATTTGACCTGGTTGCAGAAATGCGCA 2391
      |||||
1605 GlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetVa 1621
      |||||
2392 GGCAGACACTCAGAAAGCAATCATGTGAAGCTGCACAGCTGTCTAGT 2441
      |||||
1621 LHisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuLeuGlnAspGlnA 1638
      |||||
```

```
2442 CCAGCTCAGACGACTTGTGCTGAAATATTTGACATCTGCAGACCGGA 2491
      |||||
1638 rHisLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnVal 1654
      |||||
2492 AATATCTTCTGCGGATGATGTACATTTCCAGAAATTTTTCATCTAATGTT 2541
      |||||
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGluG 1671
      |||||
2542 TTGAGAAGATCTCGCGCTCAGATGATGTGTATCTCCAGATGAAGAAAG 2591
      |||||
1671 yPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuG 1688
      |||||
2592 TATCTGCTCTGGAATAATCTTATCTAGTCAGACACTTGTGGGATTAATG 2641
      |||||
1688 LcGlnAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsn 1704
      |||||
2642 AACAAACAGCTGCTTCTCTCATGCGCGCATGATGATGAACATTAAT 2691
      |||||
1705 GluValTyrLysAsnLeuIleProIleLeuGluAlaHisAspArgLys 1721
      |||||
2692 GAAGTTTACAAAGTACTTATCTTATTCATGAAGCTAATCGGAGATCAAA 2741
      |||||
1721 sLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIleM 1738
      |||||
2742 GAACATATCCACAAATCATGTGTAACCTTCAAGACATTCACGAAAAATG 2791
      |||||
1738 ethHisGlnSerSerGlyTrpGluArgValPheGlyThrTyrPheArgVal 1754
      |||||
2792 TTCATAGAGTACTGCTGCGGAGCGGATGTTGGCCCATATTTGCTGTT 2841
      |||||
1755 GlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValTy 1771
      |||||
2842 GGTTTTATAGGAACCAAGTTCCGGGATTTGGATGACCAAGAAATTTGTTTA 2891
      |||||
1771 rLysGluProSerIleThrLysLeuAlaGluIleSerHisArgGluGluG 1788
      |||||
2892 CAGAGAGCTCGCATATACCAAACTTGCAGAGATATCTCACAGATTTGGAG 2941
      |||||
1788 LuPheTyrThrGluArgPheGlyAspAspValValGluIleLeuLysAsp 1804
      |||||
2942 GATTTACAGGAGAAAGATTTGGAGAGATGCTGTTGAAGTATCAAAAGC 2991
      |||||
1805 SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG 1821
      |||||
2992 TCTAAATCCGTGACAAAGTGAATTAATCTTACAAAGGCAATATATCA 3041
      |||||
1821 nIleThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArg 1838
      |||||
3042 GATTTACTATGTGGAGCCATCTTGTGACACATATGAGAAAGGACAGAA 3091
      |||||
1838 alThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr 1854
      |||||
3092 TCACCATATTTGCACAAAAATTTACAAATCTTCGTCGATCATGACGTACA 3141
      |||||
1855 ProPheThrProAspGlyArgAlaHisGlyGluLeuProGluGlnHisLy 1871
      |||||
3142 CCGTTTACTTGTAGATGGCCGTCGCCATGGGGAACCTTCAAGAACATTTDA 3191
      |||||
1871 sArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrA 1888
      |||||
3192 AAGGAGACCATTTGTGACTGCTCATGCTTCTTATATATTAAGAAACA 3241
      |||||
1888 rGluArgValCysHisArgGluGluThrValLeuThrProValGluVal 1904
      |||||
3242 GGGTCAATGTCACTATAAAGAAAGATCATTTAACCAAAATGAATTT 3291
      |||||
1905 AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheAlaThrG 1921
      |||||
3292 GCTATTTGAGACATCAGAAAAGACACAGAGAGTTGCAATTTGCAACACA 3341
      |||||
1921 uGlnAspProProAspAlaLysMetLeuGlnMetValLeuGlnGlySerV 1938
      |||||
3342 TCAGATTCGCCGAGACCCCAAAATGCTTCAAGATGTACTCCAGAGGATCTG 3391
      |||||
```



```

1571 sGInGUaSPProGluMeLeuLleAsPLeuMeTyraGllAaNG 1588
|||||
2293 CCAGGAGGATCCGAAATGTTGATCTAAATGACAAATGGCCAAAG 2342
1588 lYrGInGUaSPProAsPLeuArGLeuThrTriPLeuGlnAsnMeLa 1604
|||||
2343 GTTACAGACCTCTCCAGATCTGGATGACCTGGTGGAGCAATGGCA 2392
1605 gLYLysHsAlaGluDeuGlnYAsnHsAlaGluAlaGlnCysMeVa 1621
|||||
2393 GGCAGACCTCAGAACGAGCAATCATCTGAAGCTGCACAGTGTCTACT 2442
1621 lHsAlaAlaAlaLeuValAlaGluTyrlLeuAlaLeuLeuGlnAspGlnA 1638
|||||
2443 CCAGCTGAGAGCACTGTGCTGAATATTGAGCATGTCGAGGAGCCGGA 2492
1638 rGhIsLeuProValIsIcYsValSerPheGlnAsnLleSerAsnVal 1654
|||||
2493 AATATCTCTCTGGSGATGTGTACATTTCAAAATTTTCACTAATGTT 2542
1655 LeuGlnUSeRAlAlleSeRAsPAlleUSeRProAsPGLuGlnI 1671
|||||
2543 TTGAGAAGATCTGCGTCTCAGATGATGTATCTCCAGATGAAGAAG 2592
1671 yPheCysSeRgLYLysHsPheThrGlnUleuGlnYLeuValIsIcYleuG 1688
|||||
2593 TATCTGCTCTGGAATACTTACTGTAGTCAGAGACTGTGGGATTAATCGG 2642
1688 lUGlnAlaAlaGlyrPheThrMetGlyLeuTyrgLlnAlaValAsn 1704
|||||
2643 AACAAACAGACTGCTCTCTCTATGCTGCGCATGTATGAACAGTATAT 2692
1705 GlUValTyLysAsnLeuLleProLleLeuGlnAlaHsArgAsPTrLy 1721
|||||
2693 GAAGTTTAAAGTAACTTATCTTATCTTAATGAACATAATCGGGAGCAAA 2742
1721 sLYsLeuAlaAlaValHsGlyLysLeuGlnGlnAlaPheThrLysLleM 1738
|||||
2743 GAACATATCCACATATCATGTGTAACCTCAAGAAAGCATTCACAAATGG 2792
1738 eThIsGlnSeRSeRgLYrPGluArGValAlPheGlyThrTyrrPheArGVal 1754
|||||
2793 TTCATCAGAGTACTGCTGCGGAGCGGATGTTGGCACCTATTTCGTCT 2842
1755 GlyPheTyGlyAlaHsPheGlyAsPLeuAsPGLuGlnPheValTy 1771
|||||
2843 GGTTTTATGGAACCAAGTTCGGGATTTGGATGAACAAGATTGTTTA 2892
1771 rLYsGluPProSeRlLeThrLysLeuAlaGlnLleSeRHisArgLeuGlnG 1788
|||||
2893 CAAGGAGCTTGCAATTAACCAACTTGCAGAGATCTCTCAAGATTGGAGG 2942
1788 lUPheTyThrGlnArGpHeGlyAsPAsPValValGlnLleLleLysAsP 1804
|||||
2943 GATTTTACGAGAAAGATTGGAGAGATGTGGTGAATTAATCAAGAC 2992
1805 SeTyrrProValAsPlySeRlyLeuAsPSeRglnLysAlaTyrlLleG 1821
|||||
2993 TCTAAATCCGTAGACCAAGTGAATAGATCTCTAACCAAGCATTAATCA 3042
1821 nIleThrTyValGlnUProTyrrPheAsPThrTyrgLlnUleuLysAsPArG 1838
|||||
3043 GATTACCTATGTGGAGCCATACTTGTACACATATGATGAAGAGCAGAA 3092
1838 aLThrTyrrPheAsPArGAsnTyrgLYLeuArGThrPheUSeRPhCyThr 1854
|||||
3093 TCACCTATTTCCGCAAAATTAATCTCGTCGATCTCATGTACTGTACA 3142
1855 ProPheThrProAsPGLyArGAlaHsGlyLleUleuProGlnGlnHsLy 1871
|||||
3143 CCTTACTTATGATGGCCGTGCCATGGGGAACCTTCAATGAACAATTCAA 3192

```

```

1871 sArgLYsThrLeuLeuSeRThrAsPHisAlaPheProTyrlLleYsThra 1888
|||||
3193 AAGGAAGACCATCTCTGACTACCTCATGCTTCCTTATATTAATAACAA 3242
1888 rGllleArGValCysHsArgGlnUthrValLeuThrProValGlnVal 1904
|||||
3243 GGGTCATGTCACTCATTAAGAAAGAGATCATCTTAACCAATGAAGTT 3292
1905 AlAlleGlnAsPMeGlnLYsLYsThrArgGlnUleuValAlaPheArThrG 1921
|||||
3293 GCTATTGGAGACATGCCAAGAAAGACAGAGAGTTGGCATTTGGCAACACA 3342
1921 uGlnAsPProAsPAlAlYsMeLleuGlnMeTValLeuGlnGlySeRv 1938
|||||
3343 TCAGATATCCCGCAGACCCCAAAATGCTTCAGATGTACTCCAGGAGACTG 3392
1938 aLGlyProThrValAsnGlnGlyProLeuGlnUValAlaGlnValPheLeu 1954
|||||
3393 TAGGCACACAGCTGATATCAGGGGCTTTGGAAAGTTGCCAGGTTTCTGT 3442
1955 AlAGlnLleProGlnAsPProLYsLeuPheArGHisHsAsnLYsLeuAr 1971
|||||
3443 TCTGAATATCACTAGTGAACCAAGCTCTTCAGACATATATAAATGCGG 3492
1971 gLeuCYsPheLYsAsPpHeCYsLYsCYsGlnAsPAlaLeuArGlySA 1988
|||||
3493 ACTCTGCTTAAAGATTTTACTTAAAGGTGTGAAGATGCTTAAAGAAAA 3542
1988 snLYsAlaLeuLleGlyProAsPGLnLYsGlnUtyrHsArgGlnUleuGln 2004
|||||
3543 ATGAAGCTTATATTGGCGGCTTCAAAAGAGTATCAAAAGGAAATG... 3589
2005 Arg.AsnTyrcYsArGleuArGlnAlaLeuGln.ProLeuLeuThrGln 2020
|||||
3590 GCGAAACATATCTTGCCCTTAAAGAGGCCCTACAGCCCT...AGATCACAG 3636
2021 ArgLeuProGlnLeuMeTAlaProThr.....ProProGlyLeuArgAs 2035
|||||
3637 AAGAGCTCTGATTTATCCAGCCAGTATGCTTGTCCCTGGCCACAGAGA 3686
2035 nSeRLeuAsnArGAlaSeRPhArGlyAlaAsPLeu 2047
|||||
3687 TTCCTTCACTGAAATGAGCTTTGCCAAATATGATCTC 3723
seq_name: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH41952
seq_documentation_block:
ID AAH41952 standard; DNA; 3173 BP.
XX
AC AAH41952:
XX
DT 30-AUG-2001 (first entry)
XX
DE CLASP-3 nucleotide fragment.
XX
KW Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;
cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
antitumor; anti-inflammatory; antihemorrhagic; antianesthetic; dermatological; uteropathic;
ophthalmological; antithrombotic; nephrotoxic; antihypertensive; antidiabetic;
neuroprotective; antistaphylococcal; antibacterial; antisense therapy;
gene therapy; chromosome 1p31.1; ds.
KM
XX
OS unidentified.
XX
PN WO200142297-A2.
XX
PD 14-JUN-2001.
XX
PE 13-DEC-2000; 2000MO-US34171.
XX
PF 13-DEC-1999; 99US-0170453.
XX
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.

```


1500 AlaArgValLysMetGlnValThrMetSerLeuSerLeuValGlyTh 1516
|||||
1186 GCCAGGTTAAATGACAGTACCAATGTCTCATCTCTGGGGGAC 1235
1516 rThnGlnAspHeserGlnGlnHisLeuArgArgSerLeuLysThrIleL 1533
|||||
1236 ATCTCAGAAATTTATGAGAAATCTTAAGACGTTCTTAAGACTATAT 1285
1533 eutThrTyrAlaGlnGlnLysMetGlyLeuArgAspSerThrPheAlaGln 1549
|||||
1286 TGACATATGCTGAGAGAGATCTGGAATTTGAGGGAACAACATTTCCGAT 1335
1550 GlnValGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrVa 1566
|||||
1336 CAGGCCAGATCGTGGTTTTCATCTCATATGATTTCTTCTGATCTGT 1385
1566 LysMetLysGlnHisGlnLysAspProGlnMetLeuIleAspLeuMet 1583
|||||
1386 GAAATGAGAGACACCGAGAGATCTGAAATGTGATTTGATTTAATGT 1435
1583 TyrArgIleAlaArgGlyTyrGlnLysSerProAspLeuArgLeuThrTrp 1599
|||||
1436 ACAGAAATTGCCAAGAGTACCAACCTCTCAGATCTGCGATGACCTGG 1485
1600 LeuGlnAsnMetAlaGlyLysHisAlaGlnLeuGlnLysHisAlaGln 1616
|||||
1486 TTGCAGAACATGCTAGCGCACCTCAGAACGAAACATCATCTGTAAGC 1535
1616 AlaGlnLysMetValHisAlaAlaLeuValAlaGlnTyrLeuAlaL 1633
|||||
1536 TGACAGAGTCTAATCTCAGCTCAGCAGCTGTGTGCTGAATTTGACCA 1585
1633 euLeuGlnAspGlnArgHisLeuProValGlyCysValSerPheGlnAsn 1649
|||||
1586 TGCTGGAGACCGZAAATATCTCTCTCCATGCTGAACATTTGCAAT 1635
1650 ILeSerSerAsnValLeuGlnLysSerAlaIleSerAspAlaLeuLeu 1666
|||||
1636 ATTTTCATCTAATGTTTGAAGAAATCTCGGCTCTCAGATGATGGTATC 1685
1666 rProAspGlnGlnLysPheCysSerGlyLysHisPheThrGlnLeuGlyL 1683
|||||
1686 TCCAGATGAGAGAGTATCTGCTCTGGAATAATCTTTACTGAGTCAGAC 1735
1683 euValGlyLeuLeuGlnGlnAlaAlaGlyTyrPheThrMetGlyLysLeu 1699
|||||
1736 TTGTGGGATTTACTGGAACAACAGCTGCTCTCTCTATGCTGACATG 1785
1700 TyrGlnAlaValAsnGlnValTyrTyrLysAsnLeuIleProIleLeuGln 1716
|||||
1786 TATGAGGAGGTTAATGAACTTACAAAGTACTTATCTTATTCATGAAGC 1835
1716 HisArgAspTyrLysLysLeuAlaAlaValHisGlyLysLeuGlnGln 1733
|||||
1836 TTAATCGGATTCGAAGAAACCTATCCACAATTTGATAAAGCTTCAAGAG 1885
1733 LAspHesThrLysIleMetHisGlnSerSerGlyTyrPGLuArgValPheGly 1749
|||||
1886 CATTCAGCAAAATTTGTTCAATCAGAGTACGCTGGAGCGGATGTTTGC 1935
1750 ThrTyrPheArgValGlyPheTyrGlnAlaHisPheGlyAspLeuAspGln 1766
|||||
1936 ACCATATTTTCGTGTGGTTTATGGAACCAAGTTCGGGGATTTGGATGA 1985
1766 uGlnGlnPheValTyrLysGlnProSerIleThrLysLeuAlaGlnLys 1783
|||||
1986 ACAAGAAATTTGTTTACAAGAGCTGCAATTAACCAAACTTGGAGATAT 2035
1783 eThHisArgLeuGlnGlnLysPheThrGlnArgPheGlyAspAspValVal 1799
|||||
2036 CTCACAGATTTGGAGGATTTTACGGGAAAGATTTGGAAGAGATGGGTT 2085
1800 GlnIleIleLysAspSerTyrProValAspLysSerLysLeuAspSerIle 1816

|||||
2086 GAGATATTAAGAACTCTAATCTGTAGACAGAGTAAATTTGATCTTAA 2135
1816 nuLysAlaTyrIleGlnIleThrTyrValGlnProTyrPheAspThrTyrG 1833
|||||
2136 CAAGCATATATTTACATATCTACTATGTGAGCCATACTTTGACACATATG 2185
1833 LuLeuLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThr 1849
|||||
2186 AGATGAAGACAGAAATCACCATTCTTGACAAAAATTAATCATCTGTGCA 2235
1850 PheLeuPheCysThrProPheThrProAspGlyArgAlaHisGlyLys 1866
|||||
2236 TTATGTACTGTACACCTTACTTACATTTAGATGGCCGCGGCCAGGGAACT 2285
1866 uProGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPhe 1883
|||||
2286 TCATGAACATTTCAAAAGAACACATCTGTACTACGTCTCATGCTTTC 2335
1883 rOtyrIleLysThrArgIleArgValCysHisArgGlnGlnThrValIle 1899
|||||
2336 CTATATTTAAACAAAGGCTCAATCTCATTAAGAAAGATCATCTTA 2385
1900 ThrProValGlnValAlaIleGlnAspMetGlnLysLysThrArgGln 1916
|||||
2386 ACACCAATTAAGTTGCTATTTAGGACATGCAGAAAAAGACAGAGATTT 2435
1916 uAlaPheAlaThrGlnGlnAspProProAspAlaLysMetLeuGlnMet 1933
|||||
2436 GGCATTTGCAACATCAGATCCCGACAGCCCAAAATGCTTCAGATGG 2485
1933 aLeuGlnLysSerValGlyProThrValAsnGlnGlnProLeuGlnVal 1949
|||||
2486 TATCCCGAGGATCTGTAGCACCAAGTGAATCAAGGAGGCTTGGAAAT 2535
1950 AlGlnValPheLeuAlaGlnIleProGlnAspProLysLeuPheArgH 1966
|||||
2536 GCCAGGTTTTCGTGTGAATATCACTAGTACCAAGGCTTTCAGACA 2585
1966 SHsAsnLysLeuArgLeuGlnCysPheLysAspPheCysLysCysGln 1983
|||||
2586 TCATATTAATACTGCACTGCTTAAAGATTTTACTAAAGAGTGTGAAG 2635
1983 sPAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGlnTyr 1999
|||||
2636 ATGCTTAAGAAAAATTAAGGCTTAATTTGGCGGTTCAAAAGGAGTAT 2685
2000 HisArgGlnLeuGlnArgAsnTyrCysArgLeuArgGlnAlaLeuGln 2015
|||||
2686 CAAGGGAATTTG...GGGAACATATCTTCCGCTTAAAGAGGCCCTACAGC 2732
2016 ProLeuLeuThrGlnArgLeuProGlnLeuMetAlaProThr....Pr 2030
|||||
2733 CCT...AGATACAGAAAGCTCCCTCAGTTATCCAAAGCAATATGCTGT 2779
2030 oProGlyLeuArgAsnSerLeuAsnArgAlaSerPheArgLysAlaAsp 2047
|||||
2780 CCCCTGCCACAGATTTCTTCACTCGAATGAGCTTTCCGAAAAATGATC 2829
2047 eu 2047
||
2830 TC 2831
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.ABL17737
seq_documentation_block:
ID ABL17737 standard; DNA; 5589 BP.
XX
AC ABL17737;
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 4684.

[illegible]

643	glnProAargPro.....gLyThAlAlaLeuGluthProValAlcylph	656
2153	CAAAAGAAACACACAGGACTTGGCCACCGCTCAGTGGAGACTCTCTTCGGCTA	2202
656	ephTrpLleProLeuLeuGlnHISgLYATgLeuArgThGLYProPhc	673
2203	CACATGGTTGGCCTTACTGCGAAGATGGAAACATTAACTTTGGGGAATTTA	2252
673	ysLeuProValSerValAspInProProSerThySerValLeuThr	689
2253	ACCTTCGCTGTTATGGTAAATACCGCGGAAATATTACATCTTATTATACCA	2302
690	ProAspValAlAlaLeuProGlyMetArgTrpValAspGlyHisGlycyla	706
2303	CCGAAATGTTCACTCCCTCGGAATTAATGGCTGGACATACAGACCCGG	2352
706	lPheSerValGluLeuThrAlaValSerSerValHisProGluAspPro	723
2353	GTTTTCTATTATATGTAGAAGAGTCACCGCAATTCATCTTGGATTCTT	2402
723	ylcLeuAspLysPhePheThrLeuValHisValLeuGluGluGlyAlaPhe	739
2403	TCTCGATGATCTTCTCTTAATTGGCAATATCTAGACACTCGCAATATA	2452
740	ProPheArgLeuLysAspPhValLeuSerGluGlyAsnAlaGluGluGln	756
2453	CCCT.....TCCCATATTGGCAGAAATATATAGAAACAGA	2487
756	uLeuArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProGluProLeuV	773
2488	ATTAAAGAAATCTTGATGGATATGCAATATGCAAAATCGTAGGCATTTAG	2537
773	AlAlaPheSerHisHisValLeuAspLysLeuValArgLeuValIleArg	789
2538	TGAGGCATCTACCATTAATGCTTGGATAGACTTATCGAATGTGTAAGTGTG	2587
790	ProProIleLeSerGlyGlnIleValAsnLeuGlyArgGlyAlaPheG	806
2588	ACCACAAAGTTGGTGAGCAACACATGCTCTGGGATCTACGGTTTTCGA	2637
806	uAlaMetAlaHisValSerLeuValHisArgSerLeuGluAlaIaG	823
2638	AGTTCTTTGTTGGTATCATCTCTGTTGTCACTACTTAACGCATCAT	2687
823	lAspAlaArgGlyHisGlyScyProGlnLeuAlaAlaTyValHisTyAla	839
2688	ATGACCAATACGCTGGCAAAATTTGCTGATCAACATATGTCACATTTTCA	2737
840	PheArgLeuProGlyThrGluProSerLeuProAspGlyAlaProProVa	856
2738	TGTAAATTCG.....CACCCCTTTCA	2760
856	lThrValGlnAlaAlaThrLeuAlaArgGlySerGlyArgProAlaSer	873
2761	AACCAAAAGCGCATACATGCAAGTCGACGACA.....ACAGAACACT	2804
873	eutyLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAla	889
2805	TGCATTTAAGCGAATGCTACACATATAT.....GATAAATGTT	2842
890	ValAlaProGly.....SerValAsp.....	896
2843	TTTAGCAAGTGGCCGGAAGCTTGGATCGCAAAAGACTGTCATATAGTCTCT	2892
897	...AspGlyValSerArgIleLeuAlaSerLysLeuLeuHisGluGlu	912
2893	TCACTGCATGCGTCGACGGGATGCTCAAGTGGCGCTGCTTCACAGAAAT	2942
912	euaLeuGluIntPValValSerSerSerAlaValArgGluAlaIleLeu	928
2943	TAGCTTTGCAATTCGGTTGTTGGACAGGAAAGCAGCTGACTTACCATCTG	2992


```

4660 ..... 4660
1534 rTyrAlaGluGluAspMetGlyLeuArgSerThrPheAlaGluGlnV 1551
4660 ..... 4660
1551 alGlnAspLeuMetPheAsnLeuHisMetIleLeuThrAspThrValLys 1567
4660 ..... 4660
1568 MetLysGlnHisGlnGluAspProGlnMetLeuIleAspLeuMetTyrAr 1584
4660 ..... 4660
1584 glLeuAlaArgGlyTyrGlnGlySerProAspLeuArgLeuThrTrpLeug 1601
4660 ..... 4660
1601 lnsAsnMetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAla 1617
4660 ..... 4660
1618 GlnCysMetValHisAlaAlaAlaLeuValAlaGluTyrLeuAlaLeuLe 1634
4660 ..... 4660
1634 uGluAspGlnArgHisLeuProValGlyCysValSerPheGlnAsnIles 1651
4660 ..... 4660
1651 erSerAsnValLeuGluSerAlaIleSerAspAspIleLeuSerPro 1667
4660 ..... 4660
1668 AspGluGluGlyPheCysSerGlyLysHisPheThrGluLeuGlyLeuVa 1684
4660 ..... 4660
1684 lGlyLeuLeuGluGlnAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrG 1701
4660 ..... 4660
1701 lvalAlaValAsnGluValTyrLysAsnLeuIleProIleLeuGluAlaHis 1717
4660 ..... 4660
1718 ArgAspTyrLysLysLeuAlaAlaValHisGlyLysLeuGlnGluAlaPh 1734
4660 ..... 4660
1734 eThrLysIleMetHisGlnSerSerGlyTyrGluArgValPheGlyThrT 1751
4660 ..... 4660
1751 yrpHeArgValGlyPheTyrGlyAlaHisPheGlyAspLeuAspGluGln 1767
4660 ..... 4660
1768 GluPheValTyrLysGluProSerIleThrLysLeuAlaGluIleSerHi 1784
4660 ..... 4660
1784 sArgLeuGluGluPheTyrThrGluArgPheGlyAspAspValAlaGluT 1801
4661 ..... 4660
1801 ..... 4660
1801 leileLysAspSerTyrProValAspLysSerLysLeuAspSerGlnLys 1817
4698 TCATTAAAGATCCCAATACCGTTGATATTAATACCTTGATCCCGATAG 4747
1818 AlaTyrIleGlnIleThrTyrValGluProTyrPheAspThrTyrGluLe 1834
4748 GCTTACATTCATTAATCTTATGTTGAACCCCTACTTGAACATATGAAT 4797

```

```

1834 uLysAspArgValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheL 1851
1851 ..... 1851
4798 GCGTCATCGTGAGACATACTTTCAGCGGAATTTCAATATAAAGCTTTTA 4847
1851 eurPheCysThrProPheThrProAspGlyArgAlaHisGlyGluLeuPro 1867
4848 TATATGCCACGCCCTTTTACTAAATAAGGAAGCAGCATGGCGAATTAAT 4897
1868 GluGlnHisLysArgLysThrLeuLeuSerThrAspHisAlaPheProTy 1884
4898 GAGCAGTGTAAACGGAAACGATTTTGACACAGCGGATCACTTCCATTA 4947
1884 rIleLysThrArgIleArgValCysHisArgGluGluThrValLeuThrP 1901
4948 CGTAAAAACCTCGCATTTATGTTATTAAGCCAGCAAAATAGCTCGTAAC 4997
1901 roValGluValAlaIleGluAspMetGlnLysLysThrArgGluLeuAla 1917
4998 CGATTGAAGTGGCAATGAGATATTCAAAATAAATAACATTGGAAATTGGCA 5047
1918 PheAlaThrGluGlnAspProProAspAlaLysMetLeuGlnMetValLe 1934
5048 GCTGCCACGAAATCAAGAGCCAGACAGCCCAAAATATATGCAATGTACT 5097
1934 uGlnGlySerValGlyProThrValAsnGlnGlyProLeuGluValAlaG 1951
5098 GCAGGATGTATGGAAACCACTGTTAACCAAGACCGATGGAAATGGCGA 5147
1951 lvalPheLeuAlaGluIleProGluAspProLysLeu...PheArgHis 1966
5148 GTGTGTCCTTTCCAAATTTATCCGACGGAACAACGTGTACCGCAAAAC 5197
1967 HisAsnLysLeuArgLeuGlyAspPheLysAspPheCysLysCysGluAs 1983
5198 CAACAACAACCTCGGTGCGTTTCGCGAGTTTGCAGAGCTTGCTGCA 5247
1983 pAlaLeuArgLysAsnLysAlaLeuIleGlyProAspGlnLysGlyTyrH 2000
5248 TCGTTTGAATAAAGAAATCGCAATCTTAATACTTTCAGATCAAAAAGATACC 5297
2000 lAsArgLysLeuGluGluArgAsnTyrCysArgLeuArgGluAlaLeuGlnPro 2016
5298 AACGAGAACTGGAGACTAACCAACGATCGGTTCATTTGAACGATTAAGTCC 5347
2017 LeuLeuThr 2019
5348 TTTATTAAT 5356
seq_name: /SIDS1/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH43850
seq_documentation_block:
ID AAH43850 standard; cDNA; 2148 BP.
XX
AC AAH43850;
XX
DT 04-SEP-2001 (first entry)
XX
DE Preliminary human CLASP-7 encoding cDNA sequence.
XX
KW Human; CLASP-7; cadherin-like asymmetry protein; immune response;
KW neuroprotective; antidiabetic; immunosuppressive; antirheumatic;
KW antitarditic; hypotensive; anti-HIV; cytostatic; immunostimulant;
KW antihaemic; antiinflammatory; ophthalmological; neproretrophic;
KW antithyroid; antiasthmatic; antibacterial; gene therapy;
KW chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;
KW juvenile diabetes; rheumatoid arthritis; pruritic urticarial papule;
KW hypertension; Rh incompatibility; ss.
OS Homo sapiens.
XX
Key Location/Qualifiers
FH CDS 1..1932
FT

```

```
FT      /*tag= a
FT      /partial
FT      /product= "preliminary CLASP-7 protein"
FT      /note= "no start codon given"
XX
XX      MO200142295-A2.
XX
XX      14-JUN-2001.
XX
XX      13-DEC-2000. 2000MO-US34152.
XX
XX      13-DEC-1999. 99US-0170453.
XX      14-JAN-2000. 2000US-0176195.
XX      14-FEB-2000. 2000US-0182296.
XX      11-APR-2000. 2000US-0196267.
XX      11-APR-2000. 2000US-0196460.
XX      11-APR-2000. 2000US-0196527.
XX      11-APR-2000. 2000US-0196528.
XX      11-APR-2000. 2000US-0547276.
XX      13-OCT-2000. 2000US-0240503.
XX      13-OCT-2000. 2000US-0240508.
XX      13-OCT-2000. 2000US-0240539.
XX      13-OCT-2000. 2000US-0240543.
XX
XX      (ARBO-) ARBOR VITA CORP.
XX
XX      Lu P, Garman JD, Candia AF;
XX      WPI: 2001-381641/40.
XX      P-PSDB; AAB99540.
XX
XX      Novel cadherin-like asymmetry protein-7 and polynucleotides encoding
XX      the polypeptide, useful for treating autoimmune disease,
XX      hypersensitivity, preventing transplant rejection by modulating immune
XX      response -
XX
XX      Example 3; Fig 1; 151pp: English.
XX
XX      The present invention describes a human cadherin-like asymmetry protein
XX      (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
XX      nucleotide sequence (II) have activities including: neuroprotective;
XX      antidiabetic; immunosuppressive; antirheumatic; antiarthritic; anti-HIV;
XX      hypotensive; cytostatic; immunostimulant; antianemic; antineoplastic;
XX      ophthalmological; nephrotoxic; antihypertensive; antidiabetic; antineoplastic;
XX      antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
XX      useful for detecting the CLASP-7 polypeptide. (II) is useful for
XX      producing (I) by recombinant methods. (I) or its fragment are useful for
XX      inhibiting an immune response in a cell such as T cell or B cell. A
XX      pharmaceutical composition (C), comprising (I) or (II), can be useful
XX      for treating CLASP-7-mediated disease such as an autoimmune disease
XX      caused or exacerbated by increased activity of T helper cells. Autoimmune
XX      diseases which can be treated using (C) include multiple sclerosis,
XX      juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
XX      toxemia or pregnancy induced hypertension, pruritic urticarial papules
XX      and Rh incompatibility. (I) is also useful as a diagnostic reagent for
XX      immune and other disorders, since diseases characterized by
XX      overproduction or depletion of lymphocytes in blood or other organs may
XX      be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
XX      mapped to the chromosomal location 19q13.2. The present sequence
XX      represents the preliminary human CLASP-7 encoding cDNA sequence which is
XX      given in the present invention.
XX
XX      Sequence 2148 BP; 468 A; 649 C; 625 G; 406 T; 0 other.
```

alignment_scores:

```
Quality: 3326.00      Length: 643
Ratio: 5.181          Gaps: 0
Percent Similarity: 99.844      Percent Identity: 99.533
```

```
alignment_block:
US-09-736-968A-2 x AAH43850 ..
```

```
Align seg 1/1 to: AAH43850 from: 1 to: 2148
1405 LeuSerGluAlaArgGluSerValLeuGlyAlaValLeuGlyValValLeu 1421
|||||
1  CTTTCAGAAAGCCCGGAGAGCGCTTGGGGGATTTGCTGAAGGTTGCTGT 50
1421 uTySerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuA 1438
|||||
51  GTACAGCCTGGGCGAGTCCAGAGTCCCTCTTTCGAGCATGGCCCTGG 100
1438 laThrglnaArgAlaLeuValSerLysPheProGluLeuLeuPheGluGlu 1454
|||||
101  CCACCAGAGGCGCCCTTGTCTCAAGTCCCGAGCTGCTGTTCGAGAG 150
1455 AspThrGluLeuCysAlaAspLeuGlyLeuLeuLeuArgHisCysGcl 1471
|||||
151  GACAGGAGCTGTGTCCAGCTGTGCTGAGAGCTCTTACGACACTGTGG 200
1471 ySerArgIleSerThrIleArgThrHisAlaSerAlaSerLeuTyLeuL 1488
|||||
201  CAGCCGCATCAGACACATCCGACAGCAGCCGCTGCTGTACCTGC 250
1488 euMetArgGlnAsnPheGluIleGlyHisAsnPheAlaArgValLysMet 1504
|||||
251  TCATGCGACAGAACTTCGAGATCGGCCACAACTTGGCCGTGAAGATG 300
1505 GlnValThrMetSerLeuSerSerLeuValGlyThrThrGlnAsnPheS 1521
|||||
301  CAGGTACCATGTCTCTCTGCTGCTGGGGGAGACGACGAACTTCAG 350
1521 rGluGlnHisLeuArgArgSerLeuLysThrIleLeuThrTyAlaGluG 1538
|||||
351  TGAAGAGCAGCTGCGAGCGTTCATCAAAACATCTCAACATATGCTGAG 400
1538 LuAspMetGlyLeuArgAspSerThrPheAlaGlnValGlnAspLeu 1554
|||||
401  AGCAGATGGGGCTGCGGACACACCTTCGCGACAGCTGCGAGACCTG 450
1555 MetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGlnH 1571
|||||
451  AGCTTCACCTGCACATGATCTTGACGACAGCGGTAAGATGAAAGACA 500
1571 sGlnGluAspProGluMetLeuIleAspLeuMetTyArgIleAlaArgG 1588
|||||
501  CCAGAGAGACCTGAGATGCTATCATGACCTCATGACGAATTCGCCGG 550
1588 lTyThrGlnGlySerProAspLeuAlaGlyLeuThrTrpLeuGlnAsnMet 1604
|||||
551  GCTACCGAGGGCTCACCGGACCTTCGCTGACCTGCTGCAGAACATGGCC 600
1605 GlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAlaGlnCysMetVa 1621
|||||
601  GGGAGAGCAGCGGAGACTGGGACACACCGCAGCGCCGCTGATGCT 650
1621 HisAlaAlaAlaLeuValAlaGluTyLeuAlaLeuLeuGlnAspGlnA 1638
|||||
651  GCACCGCGCGCCCTCGCTGCTGAGTACCTGCCCTGCTGAGAGCACACC 700
1638 rGlnLeuProValGlyCysValSerPheGlnAsnIleSerSerVal 1654
|||||
701  GCCACCTGCCCTGGGCTGCGCTTCTTCAGAACATCTCATCCAAAGTGG 750
1655 LeuGluGluSerAlaIleSerAspAspIleLeuSerProAspGluGluG 1671
|||||
751  CTAGAGGAGTCCGCCATCTCCGAGACATCTGTGCGCGAGAGAGG 800
1671 yPheCysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuLeuG 1688
|||||
801  CTTTCGTCTCGGGAAGCACTTCACCTGAGCTGGGCTGTAGGCTGCTGG 850
1688 lGlnAlaAlaGlyLysPheThrMetGlyGlyLeuTyArgValAlaAsn 1704
|||||
851  AACAGGAGCGGCTACTTACCATGGGCGGCTTACGAGGCGGTGAAT 900
```

1705	GLUValTyrTrpAsnIleuProIleLeuGluAlaHisArgAspTyr	1721
1706	GGGCTCTCAAGAACCTCATCCCATCTGTGAAGCCACCGATCA	950
1721	slsYsleuAlaIaIaHisGlyLeuGlnGluAlaPheThrIysIle	1738
951	GAAGGTGGCCGGGTGCACGGCAACATGCAGAGGGCTTCACCAAGATCA	1000
1738	eHisGlnSerSerGlyTrpGluArgValPheGlyThrTyrPheArgVal	1754
1001	TCCACCAAGATCCGGCGGGAGCGCGTGTGGGAGATTTCCGGCGG	1050
1755	GlyPheTyrGlyAlaHisPheGlyAspLeuAspGlnGlnPheValTyr	1771
1051	GGCTTCTAGCGGCCCACTTCGGTAGCTGGATGAGCAGAGATTTGTGA	1100
1771	rLysGluProSerIleThrIysLeuAlaGluIleSerHisArgLeuGlnG	1788
1101	CAGAGGCCATCGATCAAGAGCTGGGAGAGATCTCACCGCGCTGGAG	1150
1788	IuPheTyrThrGlyArgPheGlyAspAspValGluIleIleLysAsp	1804
1151	AGTTCTACACGAGAGATTTGGCCACGACGTGTTGAGATTATCAAAAGC	1200
1805	SerTyrProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleG	1821
1201	TCTTACCCCTGTGGACAACTCCAACTTGACATCAACAAAGGCTCATCTCA	1250
1821	nIleThrTyrValGluProTyrPheAspThrTyrGlnLeuLysAspArgV	1838
1251	GATCACGATGTGGGAACCGTACTTGATACCTACGACCTCAAGGACCGGG	1300
1838	AluThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThr	1854
1301	TGACCTACTTTGACCGCAACTATGGGCTTCGACATCTCTGTCTCGAGC	1350
1855	ProPheThrProAspGlyValArgAlaHisGlyLeuProGluGlnHisLys	1871
1351	CCCTTACACCCCGAGTGGGGCGGCACACGGGGAGGTGGCCGAGCAACAA	1400
1871	sArgLysThrLeuLeuSerThrAspHisAlaPheProTyrIleLysThrA	1888
1401	GCTTAGACAGCGTGTCTACGACCGACACGCGCTTCCCTACTACTAAGACTC	1450
1888	rgLleArgValCysHisArgGlnGluThrValLeuThrProValGluVal	1904
1451	GCATCCGCTGTGGCACCGGAGGAGACGCTGCTGACGCCAGTGTGAAGTG	1500
1905	AlaIleGluAspMetGlnLysLysThrArgGluLeuAlaPheIleArgI	1921
1501	GGCATGAGAGACATGGAGAAAGACAGGGAGCGCTTGTGGCACCGGA	1550
1921	uGlnAspProProAspAlaLysMetLeuGlnIleMetValLeuGlnIleSerY	1938
1551	GCAGAGACCCACCAAGATGCTAAGATGCTACAGATGGTGCTTCAGGGCTCTG	1600
1938	AlaGlyProThrValAsnGlnGlyProLeuGluValAlaGlnValPheLeu	1954
1601	TAGGGCCCAACCGTAAACCAAGGGTCCCTCGAGGTGGCCCAAGTGTTTTAA	1650
1955	AlaGlnIleProLysProLysLeuPheArgHisHisAsnLysLeuArg	1971
1651	GCAGAGATCCCGGAAACCCCAAGCTCTTCCGGCATCAACAAATTTGGC	1700
1971	gLeuCysPheLysAspPheCysLysLysCysGlnAspAlaLeuArgLysA	1988
1701	GCTCTGCTCAAGAGACTTCTGCAGAAATATGTAGAGATCGCTCGGGAAAA	1750
1988	snLysAlaIleuIleGlyProAspArgLysGluTyrHisArgGluLeuGlu	2004
1751	ATAAGGCGCCTGATTTGGCCGGACACAAAGAGATACCACCGTACACTGGAG	1800

XX	seq_name= /SIDSL/gcdata/geneseq/geneseqn-emb1/NA2001A.DAT:AA507382
XX	seq_documentation_block:
XX	ID AA507382 standard; CDNA; 4027 BP.
XX	AA507382;
XX	26-SEP-2001 (first entry)
XX	Human DNA associated with CLASP-5 #2.
XX	Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;
XX	immunogen; antibody; autoimmune disease; rheumatoid arthritis;
XX	multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
XX	acquired immunodeficiency syndrome; AIDS; ss.
XX	Homo sapiens.
XX	Key Location/Qualifiers
XX	CDS 1..2970
XX	FT /*tag= a
XX	FT /product= "Protein as displayed in AAU04026"
XX	FT /partial
XX	FT /note= "No start codon"
XX	WO200142296-A2.
XX	14-JUN-2001.
XX	13-DEC-2000; 2000WO-US34163.
XX	13-DEC-1999; 99US-0170453.
XX	14-JAN-2000; 2000US-0176195.
XX	14-FEB-2000; 2000US-0182296.
XX	11-APR-2000; 2000US-0196267.
XX	11-APR-2000; 2000US-0196460.
XX	11-APR-2000; 2000US-0196527.
XX	11-APR-2000; 2000US-0196528.
XX	11-APR-2000; 2000US-0547276.
XX	13-OCT-2000; 2000US-0240503.
XX	13-OCT-2000; 2000US-0240508.
XX	13-OCT-2000; 2000US-0240543.
XX	13-OCT-2000; 2000US-0240539.
XX	(ARBO-) ARBOR VITA CORP.
XX	Lu P, Garman JD, Candia AF;
XX	WPI; 2001-367865/38.
XX	P-PSDB; AAU04026.
XX	CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
XX	treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
XX	Rheumatoid arthritis -
XX	Disclosure; Fig 9C; 188pp; English.
XX	The sequence encodes a protein presented as AAU04026, included in the
XX	specification which relates to a cadherin-like asymmetry protein,
XX	CLASP-5, which is a transmembrane protein of the immune system involved
XX	in the immune response. CLASP-5 polynucleotides encoding

[illegible]

142 GATGTCAGGCTCCGGCTTGGAGAGGCTTTTGGCTGGCTGGGGGAGGGGCCAG 794

```

1332 gGlnGluMetValArg.....SerArgGluArgSerProPheG 1346
      |  |||||:::|||||  :::::||||  |||  |
795 AGGGGAGATGATGCGCCGCCGGCTCCAGGGAGACGACCGATTTCCA  G 841

```

1346 1YASNPROGLUASNVALARGTPARGLYSSERVALTHRIETRPGLG1N 1362

842 GCCCTAAAGTATTCAGATGGAGAAAGAGCAGACACATTTGGCGGCAA 891

1363 ThrSeraspargValAspLysThrLysaspGluMetGluHisGluAlaLe 1379
 :::::::::::::::::::::::::::: ||::::::::::::::::::::::::::
 892 GCTAATGAGAGACTAGATAAAACAAGGCCGAGTTAGATCAAGAGACCTT 941

1379 uValGIUGLYAsnLeuAlaThrGIuAlaSerLeuValValLeuAspThrL 1396

942 GATCAGTGGCAATCTGGCTACAGAAGCACATTTAATCATCTCGATATGC 991

[illegible]

332 AGGAAACATTAATCCAGGCGAGCTCGGCTCTGGACTGTAAAGACAGCCCTG 1041
1413 LeuGlyAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSe 1429

1042 CCGGAGGCGTCTGGTGAATTCCTCTGACCTGTGATCAGAG 1091

1429 PAlaLeuPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerL 1446
|::: ::||| ||| :::||||| | |||||:::|
1092 TACCACCTACTGAGTCACTGTTGTGAACACATCCGGTGTTCATGCCA 1141

1446 yspheProGluLeuLeuPheGluGluaspThrGluLeuCysAlaAspLeu 1462

```

1142 AGTTGGAGACTTACTCTTCGAGAGAGGAGTGAACAGTGTTCGACTTA 1191

```

1 —
 2 —
 3 —
 4 —
 5 —
 6 —
 7 —
 8 —
 9 —
 10 —
 11 —
 12 —
 13 —
 14 —
 15 —
 16 —
 17 —
 18 —
 19 —
 20 —
 21 —
 22 —
 23 —
 24 —
 25 —
 26 —
 27 —
 28 —
 29 —
 30 —
 31 —
 32 —
 33 —
 34 —
 35 —
 36 —
 37 —
 38 —
 39 —
 40 —
 41 —
 42 —
 43 —
 44 —
 45 —
 46 —
 47 —
 48 —
 49 —
 50 —
 51 —
 52 —
 53 —
 54 —
 55 —
 56 —
 57 —
 58 —
 59 —
 60 —
 61 —
 62 —
 63 —
 64 —
 65 —
 66 —
 67 —
 68 —
 69 —
 70 —
 71 —
 72 —
 73 —
 74 —
 75 —
 76 —
 77 —
 78 —
 79 —
 80 —
 81 —
 82 —
 83 —
 84 —
 85 —
 86 —
 87 —
 88 —
 89 —
 90 —
 91 —
 92 —
 93 —
 94 —
 95 —
 96 —
 97 —
 98 —
 99 —
 100 —
 101 —
 102 —
 103 —
 104 —
 105 —
 106 —
 107 —
 108 —
 109 —
 110 —
 111 —
 112 —
 113 —
 114 —
 115 —
 116 —
 117 —
 118 —
 119 —
 120 —
 121 —
 122 —
 123 —
 124 —
 125 —
 126 —
 127 —
 128 —
 129 —
 130 —
 131 —
 132 —
 133 —
 134 —
 135 —
 136 —
 137 —
 138 —
 139 —
 140 —
 141 —
 142 —
 143 —
 144 —
 145 —
 146 —
 147 —
 148 —
 149 —
 150 —
 151 —
 152 —
 153 —
 154 —
 155 —
 156 —
 157 —
 158 —
 159 —
 160 —
 161 —
 162 —
 163 —
 164 —
 165 —
 166 —
 167 —
 168 —
 169 —
 170 —
 171 —
 172 —
 173 —
 174 —
 175 —
 176 —
 177 —
 178 —
 179 —
 180 —
 181 —
 182 —
 183 —
 184 —
 185 —
 186 —
 187 —
 188 —
 189 —
 190 —
 191 —
 192 —
 193 —
 194 —
 195 —
 196 —
 197 —
 198 —
 199 —
 200 —
 201 —
 202 —
 203 —
 204 —
 205 —
 206 —
 207 —
 208 —
 209 —
 210 —
 211 —
 212 —
 213 —
 214 —
 215 —
 216 —
 217 —
 218 —
 219 —
 220 —
 221 —
 222 —
 223 —
 224 —
 225 —
 226 —
 227 —
 228 —
 229 —
 230 —
 231 —
 232 —
 233 —
 234 —
 235 —
 236 —
 237 —
 238 —
 239 —
 240 —
 241 —
 242 —
 243 —
 244 —
 245 —
 246 —
 247 —
 248 —
 249 —
 250 —
 251 —
 252 —
 253 —
 254 —
 255 —
 256 —
 257 —
 258 —
 259 —
 260 —
 261 —
 262 —
 263 —
 264 —
 265 —
 266 —
 267 —
 268 —
 269 —
 270 —
 271 —
 272 —
 273 —
 274 —
 275 —
 276 —
 277 —
 278 —
 279 —
 280 —
 281 —
 282 —
 283 —
 284 —
 285 —
 286 —
 287 —
 288 —
 289 —
 290 —
 291 —
 292 —
 293 —
 294 —
 295 —
 296 —
 297 —
 298 —
 299 —
 300 —
 301 —
 302 —
 303 —
 304 —
 305 —
 306 —
 307 —
 308 —
 309 —
 310 —
 311 —
 312 —
 313 —
 314 —
 315 —
 316 —
 317 —
 318 —
 319 —
 320 —
 321 —
 322 —
 323 —
 324 —
 325 —
 326 —
 327 —
 328 —
 329 —
 330 —
 331 —
 332 —
 333 —
 334 —
 335 —
 336 —
 337 —
 338 —
 339 —
 340 —
 341 —
 342 —
 343 —
 344 —
 345 —
 346 —
 347 —
 348 —
 349 —
 350 —
 351 —
 352 —
 353 —
 354 —
 355 —
 356 —
 357 —
 358 —
 359 —
 360 —
 361 —
 362 —
 363 —
 364 —
 365 —
 366 —
 367 —
 368 —
 369 —
 370 —
 371 —
 372 —
 373 —
 374 —
 375 —
 376 —
 377 —
 378 —
 379 —
 380 —
 381 —
 382 —
 383 —
 384 —
 385 —
 386 —
 387 —
 388 —
 389 —
 390 —
 391 —
 392 —
 393 —
 394 —
 395 —
 396 —
 397 —
 398 —
 399 —
 400 —
 401 —
 402 —
 403 —
 404 —
 405 —
 406 —
 407 —
 408 —
 409 —
 410 —
 411 —
 412 —
 413 —
 414 —
 415 —
 416 —
 417 —
 418 —
 419 —
 420 —

1192 TGTACCAAGTCTGTGACCCACTGACGACGACGACGATGGATGTCAACCCGAG 1241
1479 rHis1aSer1aSerLeuTyrlLeuLeuMetArgGlnAspHeuIleG 1496

1242 CCAGGCGTGGCCACCTTACCTCCTCATGAGGTTCAGTTTGGAGCCA 1291

[illegible]

1513 LeuValGlyThrThrGlnAsnPhSerGlnGlnHisLeuAraAraSerLeu 1529

1342 TTGGTGGGAGAGCACCAGACTTTAATGAGAGCACCCTGAGAAGATCCTT 1391

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

1546 hrpheaIaGlInValIGlnAspLeuMetPheAsnLeuHisMetIleLeu 1562
|||||.....:||||| |||||..... |||||
1442 CTTTCCACCACAGGTGAGAACTTCTGTAAATCTGAATAGATCTTA 1491
1563 ThrAspThrValylsMetIysGlnHisGlnIAspProGluMetLeu1 1579
|||||.....:||||| |||||..... |||||
1492 TATGACACAGGAAATGAGGAATTCAGAGAGATCCTGAGATCTTAT 1541
1579 eAspLeuMetIylArgIleAlaArgIylTyrGlnGlySerProAspLeu 1596
:|||||.....:|||||.....:|||||.....:|||||
1542 GGATCTCATGTACAGAAATGCCAAGAGTTACAGGCAATCTCTGATCTGC 1591
1596 rgleuThrTyrLeuGlnIAsnMetAlaGlyIysHisAlaGluLeuGlyAsn 1612
|||||.....:|||||.....:|||||.....:|||||
1592 GGCTTACCTGGCTCCAGACATGGCAGAGAAACACACCAAGAAAGAGCC 1641
1613 HisAlaGluAlaAlaGlnCysMetValHisAlaAlaAlaLeuValAlaGl 1629
:|||||.....:|||||.....:|||||.....:|||||
1642 TACACGAGAGCTGCATGTGCTGGTCACGCCCTGCTTATGTCGTA 1691
1629 uTyrIleuAlaLeuTleuGlnAspGlnArgHisLeuProValGlyCysValS 1646
|||||.....:|||||.....:|||||.....:|||||
1692 GTATCTGAGCATGTGGAGAGACACAGCTACCTGCCGTGGGAGTCTCA 1741
1646 erPheGlnAsnIleSerSerAsnValLeuGlnGluSerAlaIleSerAsp 1662
|||||.....:|||||.....:|||||.....:|||||
1742 GCTTCACAAATTTCTTCCAAATGTCGAGAGAGTGTGTGTCCTGAG 1791
1663 AspIleLeuSerProAspGlnGluGlyPheCysSerGlyIysHisPheTh 1679
||| |||||.....:||||| |||||.....:|||||
1792 GACACCTGTCTCACCTGACAGAGATGGGTGTCGACGAGCCAGTACTTAC 1841
1679 rGluLeuGlyLeuValGlyLeuLeuGlnIleAlaAlaGlyTyrPheThm 1696
||| |||||.....:||||| |||||.....:|||||
1842 CGAGAGTGGCTGTGAGGCTCTCTGAGACAGGCCCGGAGGCTTCCACA 1891
1696 eGlyGlyLeuTyrGlnAlaValaIAsnGluValTyrIysAsnLeuIlePro 1712
||| |||||.....:||||| |||||.....:|||||
1892 CGGAGAGCTTATATGACAGATTAAATGAGTCTACAGCTGTCTATCCCC 1941
1713 IleLeuGlnAlaHisArgAspTyrIysIysLeuAlaAlaValHisGlyIly 1729
|||||.....:|||||.....:|||||.....:|||||
1942 ATCCTAGAAGCCGATCGAGAATTCGGAAGCTGACACTCCTACAGCA 1991
1729 sIeuGlnGluAlaPheThrIysIleMetHisGlnSerSerGlyTyrGlu 1746
|||||.....:|||||.....:|||||.....:|||||
1992 GCTGCAGAGAGCTTGCAGACATGCTTAACAAGATCAT.....AGA 2035
1746 rGValPheGlyThrTyrPheArgValGlyPheTyrGlyAlaHisPheGly 1762
||| |||||.....:||||| |||||.....:|||||
2036 GAATGTGTGGAACTCTACTCCGAGTGTCTTCTTGATCCAAATTTGGC 2085
1763 AspLeuAspGlnGlnGlnPheValTyrIysGlnProSerIleThrIysLe 1779
|||||.....:|||||.....:|||||.....:|||||
2086 GATTTGGATGAACAGAGATTGTCTACAAAGAGCTGCAGATTACCAAGCT 2135
1779 uAlaGluIleSerHisArgLeuGlnIlePheTyrThrGluArgPheGly 1796
| |||||.....:||||| |||||.....:|||||
2136 TCCTGAGATCTCATAGACTAGAGCATTTTATGCTCAATGTTTGGG 2185
1796 sPAspValValGluIleIleIysAspSerTyrProValAspIysSerIys 1812
: |||||.....:|||||.....:|||||.....:|||||
2186 CAGAAATTTGGAGAGTGAATTAAGACTCCACTCTGTGCAAAACCAAG 2235
1813 LeuAspSerIleIysAlaTyrIleGlnIleThrTyrValGlnProTyrPh 1829
|||||.....:|||||.....:|||||.....:|||||
2236 TTGGATCTTACAAAGGCTCATACAGATCACTTTGTGGAGCCCTACTT 2285
1829 eAspThrTyrGlnLeuIysAspArgValThrTyrPheAspArgAsnTyrG 1846
||| |||||.....:||||| |||||.....:|||||
2286 TGATGATGATGATGAAGAGAGGCTCATACTTGGAGAAATTTTCA 2335

1846 IyLeuArgThrPheLeuPheCysThrProPheThrProAspGlyArgAla 1862
:|||||.....:|||||.....:|||||.....:|||||
2336 ACCTCGGAGAGTTCATGATACACCACCCGTTCACTCGAGGGGGGCGCT 2385
1863 HisGlyGluLeuProGlnHisIysArgIysThrIleLeuSerThrAs 1879
:|||||.....:|||||.....:|||||.....:|||||
2386 CGGGAGAGCTGCATGACAGATACAGAGAAACACAGTCTTCAACCACTT 2435
1879 pHisAlaPheProTyrIleIysThrArgIleArgValCysHisArgGluG 1896
|||||.....:|||||.....:|||||.....:|||||
2436 GCAGGCTTCCCTTACATCAAGACACGATCAGCTCATCCAGAAAGAG 2485
1896 IuThrValLeuThrProValGluValAlaIleGlnAspMetGlnIysIys 1912
||| |||||.....:||||| |||||.....:|||||
2486 AGTTTGTTTTGACACCGATTGAAGTTGCCATTGAAGACATGAAGAGAAAG 2535
1913 ThrArgGluLeuValAlaPheAlaThrGlnIAspProProAspAlaIysMe 1929
||| |||||.....:||||| |||||.....:|||||
2536 ACCCTGCAGTTTACAGCTGTCATTAACAGAGAGCCGCTGATCCAAAGAT 2585
1929 tLeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGlnGlyP 1946
|||||.....:|||||.....:|||||.....:|||||
2586 GCTTCAGATGTGCTGCAAGGCTGTGGGAGCTACTGTAAATCAGGAGAC 2635
1946 rOleuGluValAlaGlnValPheLeuAlaGluIleProGluAspProIys 1962
|||||.....:|||||.....:|||||.....:|||||
2636 CACTGGAAGTAGCCCAAGTGTGTTTGCTGAATCTCTGCTGATCCAAA 2685
1963 LeuPheArgHisHisAsnIysLeuArgLeuCysPheIysAspPheCysIly 1979
|||||.....:|||||.....:|||||.....:|||||
2686 CTCATGTGACATCACACAAAGTTGAGGTTATGCTTTAAGGAATTCATCAT 2735
1979 sIysCysGlnAspAlaLeuArgIysAsnIysAlaLeuIleGlyProAspG 1996
:|||||.....:|||||.....:|||||.....:|||||
2736 GAGATGTGCTGCAAGCTGTAGAGAAACAAAGCCGTCTATCATCGGAGACC 2785
1996 tIlyIysGluTyrHisArgGluLeuGluIysArgAsnTyrCysArgLeuArgGlu 2012
|||||.....:|||||.....:|||||.....:|||||
2786 AGAGGAATATCACAGGAACTCAAAAAGACTATACAAAGCTTAAAGAG 2835
2013 AlaleuGlnProLeuThrGlnArgLeuProGlnLeuMetAlaPro.. 2028
|||||.....:|||||.....:|||||.....:|||||
2836 AACCTCAGAGCCCAATGATGAGCGGAGAAATTCAGAACTGTACAGCCAAAT 2885
2029ThrProProGlyLeuArgAsnSerLeuAsnArgAlaSerPhe 2043
: |||||.....:|||||.....:|||||.....:|||||
2886 ATTCAGAGTTGAGAGTCAAAAGAGGAGCTCCTTCACAGATCTAGTTCA 2935
2043 rGlyIysAlaAsp 2046
|||||.....:|||||.....:|||||.....:|||||
2936 GGAAATGTGAA 2946
seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.AAS08334
seq_documentation_block:
ID AAS08334 standard; cdna: 6454 BP.
XX AAS08334;
XX
XX
DT 23-Oct-2001 (first entry)
XX
DE Human cDNA encoding CLASP-4.
XX
XX Human; CLASP-4; cadherin-like asymmetry protein-4; autoimmune disease;
KW ataxia telangiectasia; human immunodeficiency virus infection;
KW inflammatory disease; rheumatoid arthritis; multiple sclerosis;
KW diabetes mellitus; immune disorder; guillain-Barre syndrome;
KW severe combined immunodeficiency; allergic reaction; asthma;
KW immunogen; antibody; ss.
XX
XX Homo sapiens.
OS
XX
XX
FH Key Location/Qualifiers

1316	CCTGAATTTTCTTAGTTCGCCAGAAATTGAAAAGTACTA...CAGGGAAA	1362
350	pIleSerGIuCyScySGluProTyMetValLLeuLYSGluValAspPTha	367
1363	CATTACACACTGTGCAGAACCTAT...ATCAAAAATTCTGATCCAG	1406
367	IaLYsAnLYSGluLYSLeuLIuLYSLeuArGLeuLIaIaGLuGLInPhe	383
1407	TAAAGCAGGCCAGAGGATGACAGG...ACAGCTAAACAAAGTC	1447
384	CysThArGLeuGLYArGTyArGLmeProHeaLArPThArIaValNI	400
1448	TGTAAGCCGCTTGACATACAGAAATGCCCTTGCTGGCTGCCAGACC	1497
400	sLeuLIaAnLIeValSerSerIaGLyGLInLeuAsPArAspSerAsps	417
1498	CATT...TTCAAGAATACCAAGAGCGCTCTTGATCTGGAT...1534	1534
417	erGLuGLuArGrProLaArPThArAsPArGrArGrGLyPro	433
15341534	1534
434	GIaSPArGLaSerSerLYaSPaSPalacysSerPheSerGLyPheAr	450
1535GGAGATTTTC1545	1545
450	gProLaThInLeuThValThAsnPhePheLYSGluGLInLaGLuArGL	467
1546	TCCT...CTGTATTAACMACAGACGTACAGAC1574	1574
467	eUSeRAsPGLaSPLeuPheLYSPhelEuLIaSPmeTArGrArProSer	483
1575	TTTCAAGTGAAGCATTTCCAGTTCCTGCAGAAATATAGAACCCAGAA	1624
484	SeRLeuLeuArGrGrLeuArGrProValThrIaGLInLeuLYSLeaPrl	500
1625	AAG...ACCAACTGCAGATTATTTCTCTGGCAGCTAAACATCACAGT	1668
500	eSeRProLaLProGLaSPLeuProLIaSPheCySeuSerProGLeULeuNI	517
1669	AGAAATGTTGCTCGTGATTTATTCAAATGGATGTATCTCTTCATATGTGC	1718
517	ISLeuLYSProLYrProAsPProArGLyArGrProThrLYSGLIaLeu	533
1719	CTTGAAGCCCTTTTGAAAGAAATTGCCAAAAATATTACTGTGGAGCTTGA	1768
534	GIuPhe...ProLaArGrGLuValLYaLa...ProLIaSPheSerTyAr	548
1769	GAGTTTGTCACAAATATGCAAAATATGTGTATCCATTACTATTATTTCAA	1818
548	gaSIuLeuLeuTyValTYrProLIaSerLIeuaSPheSerSerGrGLIn	564
1819	AAACCAATCTGTATGTATATCCCTGCAGAAATTAATAAAGATATGCCAGAAA	1868
565GIySeRValIaGrAsnLeuLIaValArGLTyMet...Thr	578
1869	CATTGGCAAGCGCAGAGAACATTTGCAGCTGTGTGGAAATCCGGGATTCA	1918
579	GLyGLuAsPProSeRGLInaLeuProValIIePheGLyLYSSeRSeCy	595
1919	GATCAAAAGTAGCCGTAGTGGCCCTTAAGAGTATATTATGGAAGAACTGCAGG	1968
595	sSeRGIuPheThArGLInLaPheThrProValIaLYrNIaSnLYSS	612
1969	GTCGTGTTTACCAAAATGCTTATGCTGTGTGTCGCATACAAACAAA	2018
612	erProGLuPheTYrGLuGLInLeuLYSLeuNISeuProLaLYsValThr	628
2019	ATCCAGAGTTCTATGAAGATTAATAATGACCTTCCATTCACCTCAT	2068
629	GIuLYSNIaSHsLeuLeuPheThrPheTYrNIaSVaLYSeRSeCyGLIn	643
2069	CAAAACATCATTTGCTTTTCACTTTTATCATGTAGTGTGGAATTTAA	2118

```

644 .....ProArgProGlyThrAlaLeuGluThrProValGlyP 656
      ::::: ::::: ::::: ::::: :::::
2119 CACAAAGGAGACACCAAAAGCAGACAGCTGAACTCCAGTTGGGT 2168
656 hehThrPLeuLeuGluHisGlyArgLeuArgThrGlyProHe 672
      ::::: ::::: ::::: ::::: :::::
2169 TTGCGGGGTACCTTGTGAAAAGATGAGAAATCATCATTTGAGCAG 2218
673 CysLeuProValSerValAspGlnProPro..... 682
      ::::: ::::: ::::: ::::: :::::
2219 CAGCTGCCAGTTTCCGCCAATCTTCCCGAGCTACTGAATCTCAATGA 2268
683 ...ProSerTyrSerValLeuThrProAspValAlaLeuProGlyMetA 698
      ::::: ::::: ::::: ::::: :::::
2269 TGCAGAAATCCAGAAAGCAATGTAACGTGATATT.....A 2303
698 rgTrpValAspGlyHisLysGlyValAlaPheSerValGluLeuThrAlaVal 714
      ::::: ::::: ::::: ::::: :::::
2304 AATGGGTAGATGTCGCAAAAGCCTTGTGTGAAGTTTAAAGCCACTTAGAA 2353
715 SerSerValHisProGlnAspProTyrLeuAspLysPhePheThrLeuVal 731
      ::::: ::::: ::::: ::::: :::::
2354 TCTACCATTTACACTCAAGATCTGCATGTCGCAAAATCTTCCATCATG 2403
731 HisValLeuGluGluGlyAlaPheProPheArgLeuLysAspThrValL 748
      ::::: ::::: ::::: ::::: :::::
2404 CCAGCTGATTCAGTCAGGCTCG..... 2425
748 euSerGluGlyAsnValGluGlnLeuArgAlaSerLeuAlaLeu 764
      ::::: ::::: ::::: ::::: :::::
2426 .....AAGAAGTTCAGGAGGAGCTCATTAATATTAAAGTTTG 2467
765 ArgLeuAlaSerProGluProLeuValAlaPheSerHisValLeuAs 781
      ::::: ::::: ::::: ::::: :::::
2468 CATGCCATGAGATCCCAAGTCATGATACAGTTCTACCTGTAATCTTAT 2517
781 PheLeuValArgLeuValIleArgProProIleLysSerGlyGlnLeu 798
      ::::: ::::: ::::: ::::: :::::
2518 GCAACTCTCCGAGTCTCACAATAATGACCCATGAAGATGACCTTCTTA 2567
798 AlaLeuGlyArgGlyAlaPheGluAlaMetAlaHisValSerLeu 814
      ::::: ::::: ::::: ::::: :::::
2568 TCAC.....TGCACCATGGTCTCTTACATATTGTCACAAAG 2605
815 ValHisArgSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProG 831
      ::::: ::::: ::::: ::::: :::::
2606 TGCATGAA.....GAAGGCTTGATAGTTA 2631
831 nLeuAlaAlaTyrValHisTyrAlaPheArgLeuProGlyThrGluPro 848
      ::::: ::::: ::::: ::::: :::::
2632 TCTAAGATTCATTCATAAGTAATAGCTTCCGA.....CCTGAAAACCGA 2675
848 eLeuProAspGlyAlaProProValThrValGlnAlaAlaThrLeuAla 864
      ::::: ::::: ::::: ::::: :::::
2676 GTGCTCTCTCAG..... 2686
865 ArgGlySerGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerI 881
      ::::: ::::: ::::: ::::: :::::
2686 ..... 2686
881 eSerSerSerAsnProAspLeuAlaValAlaProGlySerValAspAspG 898
      ::::: ::::: ::::: ::::: :::::
2686 ..... 2686
898 LuValSerArgIleLeuAlaSerLysLeuHisGluGlnLeuAlaLeu 914
      ::::: ::::: ::::: ::::: :::::
2687 .....GCCACCTGATACATGAACCCGCTACT 2716
915 GlnTrp.....ValValSerSerSerAla.....ValArgGluAl 926
      ::::: ::::: ::::: ::::: :::::
2717 ACGATATAGCAATATTATGAACAGTCTCGAGATTTTATCAATAAACAA 2766

```

```

926 aLeLeuGlnHisAlaTrpPhePhePheGlnLeuMetValLysSerMetA 943
      ::::: ::::: ::::: ::::: :::::
2767 ATTCTTAAGTACTCATGAGTTTCTTTGAAATATGCAAAATGCAATGG 2816
943 ILeuHisLeuLeuGlyGlnArgLeuAspThrProGlyLysLeuArg 959
      ::::: ::::: ::::: ::::: :::::
2817 CCACATGACTGTGTCGAGAGACATTAAGATTAAAGTTCCCGCAGGCCAGAG 2866
960 PheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySerValG 976
      ::::: ::::: ::::: ::::: :::::
2867 TTTCCGAGACATATGATCATGTCTTACATTCACCTGCTTCTTCCAAAT 2916
976 YLeuGluValIleThrArgValHisLysAspValGluLeuAlaGlnHis 993
      ::::: ::::: ::::: ::::: :::::
2917 TCCCGCATGACTATTCGTTATGCGGAGATTCGCCGATGACTCCAGAAATG 2966
993 euAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeuValAsp 1009
      ::::: ::::: ::::: ::::: :::::
2967 TGAATCTATAGTTGGGTAGCTCTGTAAGCGCGTGTTCACACTAATGAT 3016
1010 ArgGlyPheValPheSerLeuValArgAlaHisTyrLysGlnValAlaTh 1026
      ::::: ::::: ::::: ::::: :::::
3017 AGAGGATTTATTTCAATTATTAATGAATGACTATATATGTGATTC.... 3061
1026 rArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeuArgMetG 1043
      ::::: ::::: ::::: ::::: :::::
3062 .....AGCCCAAGATCTTAAGTCTGGCTGAATACAACTTG 3101
1043 LuPheThrArgIleLeuCysSerHisGlnHisTyrValThrLeuAsnLeu 1059
      ::::: ::::: ::::: ::::: :::::
3102 AATTCTGCAMACAAATTTGCANATCAGAACATTAATCTCTGTAACCTTG 3151
1060 ProCysCysProLeuSerProProAlaSerProSerProSerValSerSe 1076
      ::::: ::::: ::::: ::::: :::::
3152 CCAATG.....GCATTTGCCAAAACCTTAATCTGACGG 3183
1076 rThrThrSerGlnSerSerThrPheSerSerGlnAlaProAspProLysV 1093
      ::::: ::::: ::::: ::::: :::::
3184 GGTTCAAATGTCAAATCTTGAATACGT..... 3211
1093 aThrSerMetPheGluLeuSerGlyProPheArgGlnHisPheLeu 1109
      ::::: ::::: ::::: ::::: :::::
3212 .....TTATCAGATGAGTATTCAGACATCACTCTTCTG 3244
1110 AlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGluProGluAlaG 1126
      ::::: ::::: ::::: ::::: :::::
3245 GTTGGTCTACTCTGAGGGAACCTTCATGCTCTTCCAGGACATTTATGA 3294
1126 uGlyAlaPheLeuLeuHisLysLysAlaIleSerAlaValHisSerLeu 1143
      ::::: ::::: ::::: ::::: :::::
3295 G.....ATCAGATATACAGCTATCTCTGTATTAAGAATCTTT 3332
1143 euCysGlyHisAspThrAspProArgTyrAlaGluAlaThrValLysAla 1159
      ::::: ::::: ::::: ::::: :::::
3333 TGATTAACATGCAATTTGACACAGATACAGCACAGAACCAACCAACCC 3382
1160 ArgValAlaGluLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLe 1176
      ::::: ::::: ::::: ::::: :::::
3383 AAAATAGCACAATGTACTCCCTTGTGTGACTACTTTGGAANAATAT 3432
1176 uProArgLeu..... 1179
      ::::: ::::: ::::: ::::: :::::
3433 ACAGGATTAAGCAGTCCAGATACCTTGATTCTTGTGCAGCCATGCTTA 3482
1180 .....HisAspPheAlaGluGly.....ProGlyGln 1188
      ::::: ::::: ::::: ::::: :::::
3483 ATTTCGCATCCAGAGATGAGTTTCCATGTGGCTTTACTTCCACCGCAAT 3532
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThr..... 1200
      ::::: ::::: ::::: ::::: :::::
3533 AGAGGAAATCTGACACT.....GACAAAGACACCCCTTATGGGTCTT 3576
1201 ...GluGlyGluGly.....AspIleAlaGlyThrIleAsnP 1212

```

```

3577 TCMAAATGGACATGGAAATTAAGAGAGAGAGATTCAAGAGATTCCCTCATCC 3626
1212 roserValAlaMetAlaIleAlaIglyPro..... 1222
3627 CAGAGAGACA.....ACAGATTCCAGATCAGGCAACCTGGT 3667
1223 .....LeuAlaProIlySerArgAlaSerIleSerInglyProP 1236
3668 GAAAAATACCCGACAGAGTTCTCAAGAGAGTAGTATATCCAGTATAC.. 3715
1236 othraIaSerArgAlaGlyCysAlaLeuSerAlaIuSerSerArgThrL 1253
3716 .....CGCGTGGATCAGATGAAATACGAAGACC 3743
1253 enuLeuIaCysValIleuThrPValIleuLysAsn...ThrGluProAlaLeu 1268
3744 TCGGTATGTGCTACCGTATATATAGTAAATATGATTTCCAGAAATATCTTC 3793
1269 LeuGlnArgThrAlaThrAspLeuThrLeuProGlnLeuGlyArgLeuLe 1285
3794 TTAACCTACTGC...AATAAGATATACACTCAGAGCTCAATAACATTCT 3840
1285 uasPLeuLeuThrCysCysLeuAlaAlaPheGluThrIleGlyAlaArgG 1302
3841 TATACTTTTGAAGATATGCTTTGTTCACTTATAGATATATGGGAAAATA 3890
1302 laPheGluArgIleAsnSerIleuThrPheLys.....SerLeuasp 1316
3891 ACATACACAGGTCATGATGCTGCGTCAAAACACTTGGGAATAC 3940
1317 MetLysAlaArgLeuGluAlaIleLeuGlyThrIleGlyAlaArgG 1333
3941 CGAAATACCGAAACACATGCTCTCTTGGAAACAGATCAGAGATATACA 3990
1333 ngluMetValArgArg...SerArgIlaArgSerProPheGlyAsnProG 1349
3991 GGCGCGCTTCAGACATTAGTACGCTAGAAAGTTCATT..... 4030
1349 IuAsnValArgThrArgLysSerValThrHisTrpLysGlnThrSerasp 1365
4031 .....ACACTTAATCAAGTTCTTACAA.... 4054
1366 ArgValAspLysThrLysAspIleuMetGluHisGluAlaValGluG 1382
4055 .....ACGTAAACCAACAATTTCCACCGAGCACTTGTGAAGG 4092
1382 yAsnLeuAlaThrGluAlaSerLeuValValLeuaspThrLeuGluThr 1389
4093 CAATACAGCTACGAAGTTCCCTTAAGATACATACACATATCATTTT 4142
1399 IeValGln.....ThrValMetLeuSerGluAlaArgIuSer 1411
4143 TCACCTAGGCTTCACAGACCAACATTTTAAATGAATGAGCGCATACCCA 4192
1412 ValLeuGluValValLeuLysValValLeuThrYserLeuGlySerAlaG 1428
4193 TTAATGAAAAAGTGTATTATATCAATCTGCTTTTAAATAGGACA 4242
1428 nSerAlaLeuPheLeuGlnHisGlyLeuAlaAlaThrGlnIlaLeuValS 1445
4243 AUCTGAAGTGTGCTGAACATGATTTGGCTCACTGAGAGAGCTTCATCA 4292
1445 erLysPheProGluLeuLeuPheGluGluAspThrGluLeuCysAlaLasp 1461
4293 GTTAAGTTTCCTTCACGATTTTCAAAGAGAGTAAACATGTGTGTCCA 4342
1462 LeuCysLeuAlaGluLeuAlaArgHisCysGlySerAlaGlyIleSerThr 1478
4343 TTTTGTATAGAGTTTAAAGTGTGCGACATCGAAGATTAGCTCAACAG 4392
1478 gThrHisAlaSerAlaSerLeuThrLeuLeuMetArgGlnAsnPheGlu 1495

```

4393 GAATGAGACATCGACCTTTTGATCTTTTGATGAGAAACAACCTTTGAT 4442

1495 le.....glyhisaspheialaargvalysmetcInvalThrmeser 1509

4443 ATACCAAAAGGAAACCTTTTGAGAGACACATCTCAATATAATATGCT 4492

1510 LeuserSerLeuValglYthrhrGlnaspheSerGluGlnHisLeuAr 1526

4493 GTAGGCCAACTGATACCTGATAGACACTAACCGGAGGATCAAGTTTCA 4542

1526 garSerLeuLysThrIleLeuThrTyAlaGluGluaspmetGlyLeuA 1543

4543 GGAGCTTTATTCATTCATTCAAAATATTTGGCAAAATGTGCACAGACTTGA 4592

1543 rgasSerThrPheAlaGluGlnValGlnaspLeuMetPheaspLeuHis 1559

4593 AGGCACCTGCTTTTCCCGCAGAAAGTCAAAAGACTTACACACAGAACTGCC 4642

1560 MetIleLeuThrAspThrValLysMetLysGlnHisGlnaspProG1 1576

4643 ACGTTCTTATGCGCACCTGCCCAATGAAAGGAGCATGAGAAAGACCTGA 4692

1576 uMetLeuIleaspLeuMetTyAlaAlaArgGlyTyrglnGlnSerP 1593

4693 AATGCAATTTGATCTCACTCACTATGCTTACCCAGACGTCATGCACACACC 4742

1593 roAspLeuArgLeuThrTrpLeuGlnAspMetAlaGlyLysHisAlaGlu 1609

4743 CAAAGCTCAGAAAACTGCTGATGATGAGATGCCAAGATTCATGTAA 4792

1610 LeuGlyAsnHisAlaGluAlaalaGlnCysMetValHisAlaAlaLe 1626

4793 AATGGAGATTTTTCAAGAGCTCGAGTGTGTTTGTCCATGTAAGCAGCTCT 4842

1626 uValAlaGluTyLeuAlaLeuLeuGlnaspGlnArgHisLeuProValG 1643

4843 AGTTGCAGAGTTT.....CTTCATGCAGAAAATATTTTCTTAACG 4883

1643 LysCysValSerPheGlnAsnIleSerSerAsnValLeuGluGlnSerAla 1659

4884 GAGTTCAGGCTTCAAGAAAAATTTACTCCCAATATGATGAGAAAGAGACA 4933

1660 IleSerAspAspIleLeuSerProAspGluGlnGlyPheCysSerGlyLy 1676

4934 ATGAAAGAAATCTCTGGATGATGAT.....GT 4962

1676 snAspPheGluLeuGlyLeuValglYLeuLeuGlnGlnAlaAlaGlyT 1693

4963 CCAATTATGAGAAAGAGTCTGTGTGGAGTTGCTGTAACAACTGTGAGTG 5012

1693 yrPhehrMetGlyGlyLeuTyrgluAlaValaspGluValTyrgLysAsn 1709

5013 GCATTATGGAAGCGACAGACGTTATGAAATATTTTCGAAATTTCCAAATTG 5062

1710 LeuIleProIleLeuGlnLysHisArgAspTyrgLysLeuAlaAlaVala 1726

5063 ATGCTTCCAAATTATATAGAAAGAGCTGCTGAGTTGGAGAAACTTACCAAGT 5112

1726 HisGlyLysLeuGlnGluAlaAlaPheThrLysIleMetHisGlnSerSerg 1743

5113 TTTATAGAACTCTCATGAGAGCTTACACAAAATTTCTGGAAGTTATGCTAT 5142

1743 LyrTrpGluArgValIlePheGlyThrTyrgPheArgValGlyPheGlyGly 1758

5163 CAAAAAGAGACTTTTAGCAGCTTTCTTGAGGTGGCCCTTTATAGCCAA 5212

1759 AlaHisPheGlyAspLeuAspGlnGlnGlnPheValTyrgLysGluProSe 1775

5213 TCTTTTATGAGAAACAACAATGGAAGAGAGTACATCTATAAGAACAA 5262

1775 rIleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheTyrgThng 1792

5263 GCGCATGCGCTCTCGAAATTTCTCTTGAGATCTGTATAACTTATGTGTG 5312

seq_name	seq	seq_id	seq_desc
1792	lathrhecllyasprapvalvalgluilellyasprserthyproval	1792	lathrhecllyasprapvalvalgluilellyasprserthyproval
5313	aaagattgtacgagaaatgcataaataattcagattcagacaagta	5313	aaagattgtacgagaaatgcataaataattcagattcagacaagta
1809	asplyssetlyleuaspsercilysalatyrlleglnlertthyra	1809	asplyssetlyleuaspsercilysalatyrlleglnlertthyra
5363	aaagccaaagacgttgatccaaaatattcctatattacaaagttaattagt	5363	aaagccaaagacgttgatccaaaatattcctatattacaaagttaattagt
1825	lgluprtyrphesprthrytyrlyleuasprapvalthyrrhea	1825	lgluprtyrphesprthrytyrlyleuasprapvalthyrrhea
5413	gaagccttcttcttgatgcataaagaactacagaaagaaagccagatttg	5413	gaagccttcttcttgatgcataaagaactacagaaagaaagccagatttg
1842	spatgastlytyrlyleuathryrthreleuephecysthprophethpro	1842	spatgastlytyrlyleuathryrthreleuephecysthprophethpro
5463	aaagaaatcatatattatgcagagatttggttttgaggccctttacacttta	5463	aaagaaatcatatattatgcagagatttggttttgaggccctttacacttta
1859	aaaglyatgalaahisglyleuueuproblnhtslyasrlysthyrle	1859	aaaglyatgalaahisglyleuueuproblnhtslyasrlysthyrle
5513	tcagccaaanaaagagggctgtttacaaacagagccaaagccctaacat	5513	tcagccaaanaaagagggctgtttacaaacagagccaaagccctaacat
1875	uleuSerthrasphnlsalapherprotyrlyleusthryrlyleuvalc	1875	uleuSerthrasphnlsalapherprotyrlyleusthryrlyleuvalc
5563	cttgacaaacttcaactcgatttccttacctgacaaagagacttcttata	5563	cttgacaaacttcaactcgatttccttacctgacaaagagacttcttata
1892	yshiaarglyuglyuthrvalleuthrprovalgluvalalailegluasp	1892	yshiaarglyuglyuthrvalleuthrprovalgluvalalailegluasp
5613	actgtcaaacacacacatttaatttaaacacacatttgatggccacgataa	5613	actgtcaaacacacacatttaatttaaacacacatttgatggccacgataa
1909	metglnlylysthrarglyleuualaphelathrgluglnasprpropr	1909	metglnlylysthrarglyleuualaphelathrgluglnasprpropr
5663	atpaaagatpaaacacacacacacacacacacacacacacacacacac	5663	atpaaagatpaaacacacacacacacacacacacacacacacacacac
1925	oasplalysmetleuglnmetvalleuglnlyservaltyrprothry	1925	oasplalysmetleuglnmetvalleuglnlyservaltyrprothry
5713	gcacatgatttcagctcacaacttaatttcagagcgttggttctgtcagc	5713	gcacatgatttcagctcacaacttaatttcagagcgttggttctgtcagc
1942	alaspngllyproleugluvalaglnvalphelualaglulilepro	1942	alaspngllyproleugluvalaglnvalphelualaglulilepro
5763	tcnaatgctgttcatttgcataatgcacaaagcttctttaaattgacaccaa	5763	tcnaatgctgttcatttgcataatgcacaaagcttctttaaattgacaccaa
1959	gluaspprolyleuaphearghishisasnlystheargleucysphely	1959	gluaspprolyleuaphearghishisasnlystheargleucysphely
5813	gctacacagatattccacacacacacacacacacacacacacacacacac	5813	gctacacagatattccacacacacacacacacacacacacacacacacac
1975	sasprhecylystyscysgluaspralaleuathryasnllyalaleut	1975	sasprhecylystyscysgluaspralaleuathryasnllyalaleut
5863	gaattttatpacaagcatgcacacattgcctgaaactaaatgacggctaa	5863	gaattttatpacaagcatgcacacattgcctgaaactaaatgacggctaa
1992	leglyrpspqlnlysglytyrthiasrarglyleuglnatrgasnttycys	1992	leglyrpspqlnlysglytyrthiasrarglyleuglnatrgasnttycys
5913	traaagaagatcacttgatgacacacacacacacacacacacacacacac	5913	traaagaagatcacttgatgacacacacacacacacacacacacacacac
2009	argyleuathrglyalaleuglnproleuathrglnatrgleuproglnle	2009	argyleuathrglyalaleuglnproleuathrglnatrgleuproglnle
5963	gacatgataaagaatttctgcacattatpccatgacagacattattacaga	5963	gacatgataaagaatttctgcacattatpccatgacagacattattacaga
2025	umetalaprotmrproproglyleuathrgasrserleuasn	2025	umetalaprotmrproproglyleuathrgasrserleuasn
6013	ac	6013	ac
seq_name	/SIDS1/gcgcdata/geneseq/geneseq-emb1/NA2001A.DAT:AAD19118		
seq_documentation_block			
ID	AAID19118 standard; cdna; 7506 bp.		
AC	AAID19118;		
XX			
XX	18-DEC-2001 (first entry)		
XX	Angiogenesis associated human thyroid regulated transcript (hTRG) cdna.		
DE	Angiogenesis associated protein; AAP; cytosolic; cardiant; gene therapy		
XX			
XX			

```

KW ophthalmological; vulnerability; myocardial infarction; macular degeneration;
KW diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;
KW rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;
KW cancer; therapeutic; diagnostic; human; thyroid regulated transcript;
KW TRG; ss.
XX
XX Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX FH 43..6366
XX CDS /*tag= a
XX FT /product= "human thyroid regulated transcript (hTRG)"
XX FT
XX
XX W0200170808-A2.
XX
XX
XX 27-SEP-2001.
XX
XX 22-MAR-2001; 2001WO-US09609.
XX
XX 22-MAR-2000; 2000US-191134P.
XX
XX (CURA-) CURAGEN CORP.
XX PA
XX (GETH ) GENENTECH INC.
XX
XX
XX Rastelli LK, Gerritsen M;
XX
XX WPI; 2001-602775/68.
XX DR
XX P-PSDB; AAE11889.
XX DR
XX
XX Novel angiogenesis associated polypeptides and polynucleotides encoding
XX PT the polypeptides, useful for modulating angiogenesis and for treating
XX PT tumours and cancers -
XX PS
XX
XX Claim 6; Page 18-21; 159pp; English.
XX
XX The invention relates to angiogenesis associated proteins (AAP) and their
XX CC corresponding cDNA molecules, which are useful for modulating
XX CC angiogenesis. AAP proteins and nucleic acids are useful for promoting
XX CC wound healing, for example after organ transplantation, and in the
XX CC treatment of tumours, myocardial infarction, cancers, diabetic
XX CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.
XX CC AAP proteins and DNA's are useful in potential prophylactic and
XX CC therapeutic applications implicated in a variety of disorders including
XX CC those related to angiogenesis, and also in diagnostic applications.
XX CC AAP cDNA is also useful in gene therapy. The invention also relates to
XX CC a method for screening a tissue sample for tumorigenic potential. AAP
XX CC proteins are used to screen drugs or compounds that modulate AAP activity
XX CC or expression as well as treating disorders characterised by insufficient
XX CC or excessive production of AAP or production of AAP forms that have
XX CC decreased or aberrant activity compared to the wild type protein, or
XX CC modulate biological function that involve AAP. The present cDNA sequence
XX CC encodes human thyroid regulated transcript (hTRG) protein which is an
XX CC angiogenesis associated protein (AAP) of the invention. Human TRG is
XX CC involved in signal transduction between receptors and kinases. Modulation
XX CC of hTRG is useful to treat diseases related to thyroid stimulating
XX CC hormone (TSH) imbalance.
XX
XX Sequence 7506 BP; 2147 A; 1689 C; 1764 G; 1906 T; 0 other;
SQ

```

67 GCCTCGCGGAGACTTCGGAAGTTCAOCCGGCGCTGATTAAGCCGGGCAC 116
17 lAlaIaIaIaValArgLySglInValSerArgLySglSerProH 34
||||| : : : : :
117 GCGGCGCGAGCTGCGAGAGCGCTGTCTGAGTGCGCGGCTCCGTC 166
34 lSerSerArgArgYsSerSerSerLeuGlyValProLeuThrGlyVal 50
: : : : :
167 TCCTGGCAAG.....CCAAAGCTA 186
51 ValGluProLeuAspPheGluAspValLeuLeuSerArgProProAspAl 67
: : : : :
187 ATTGACCCACTGCATATATATAAATGTCATCGTCACAGAAACATCAGAT 236
67 agLuproGlyProLeuArgAspLeuValGluPheProAlaAspAspLeuG 84
: : : : :
237 CTGAAGCACTGTTTACGGGAGATGCTGCTTCCTTCATGATGACTTTC 286
84 lLeuLeuLeuGlnProArgGluCysArg.....ThrThrGluPro 97
: : : : :
287 AGACGGCCATCTGAGACGACAGGGTCGATACATATGCTCAACAGTCCT 336
98 GlyIleProLyAspGluLySLeuAspAlaGlnValArgAlaAlaValG1 114
: : : : :
337 GCGAAGCGGAGAAG3GAGACAGACAGACTGTTGTTACAGAGTCATCA 386
114 uMetYrIleGluAspTyrValIleValHisArgArgYrGlnYrLeus 131
: : : : :
387 AACCTATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
131 ertAlaIaIaYrSer..... 135
437 CAGGAGAGTTTGCAGAGCTTCGACAAAGTGCTCAAGTTGATAAAGTT 486
136 ProVal..... 137
487 CCAATTCAATGTCTATGAGTTGACGAGAGGTGACAAAGATGAGATGC 536
137 137
537 TGCTCCCTTGTGTTCCAGAAAGTGAGATCAACAAGATGCTGGCTGT 586
137 137
587 ACAAGGCAACATGAACAGTGCATGACGCTGACCATGAGTCAATTAAAG 636
137 137
637 AGACGATTTTCCACTGATTCAACTTGGCGATGATCCATATAATTGA 686
138ThrThrAspThrGlnArgGlu..... 144
687 ATTTTAAAGATCTCCAAAGAGCAACAAAGATCAATATTCTGCGAT 736
145ArgGlnLySglYeuProArgGlnValPhe 154
737 TCCTGATAGGGGTGCTTCAGGAACAAAGTCAAGCGCTTGTGCTT 786
155 Glu.....GlnAspAlaSer.....GlyAspAlaYr 163
||| : : : : :
787 GAGCTCAAGTGCAGGACAAAGTACTTATCTTGGCAGCAGACAGTGA 836
163 gSerGlyProGluAsp..... 168
837 AGTGAATAAGAGATGATGATCAATTCATAATAGATCCCTCAGCTCA 886
169SerAsnAspSerArgArgLySerGlySerProGluAsp 181
887 ACTTTGAAGCTGCAATGCAAGAAAGCAATGCGACTCTCAGCAAGAT 936
182 ThrProArgSer.....SerGlyAlaSerSerIlePheAspLeuArgAs 196
: : : : :
937 GATGACCAAGCAATTTGGAAGTTCGTGTTCC..... 969

196 nLeuAlaIaIaAspSerLeuLeuProSerLeuLeuGluArgAlaIaIaProG 213
: : : : :
970GGTTAGATAGCTACCTGCGGAACTTGCACAAAGTGCAAGAGAG 1015
213 lAspValAlaAspArgArgAsnGlnThrLeuArgArgGlnHisArgProPro 229
: : : : :
1016 CAGAAATCAAACTGAAAGTGAAGC..... 1041
230 AlaLeuLeuThrLeuYrProAlaProAspGluAspGluAlaValGluAr 246
: : : : :
1042 AGAGTCAAACTTTTATTATTGGACCAAGATGCCAAGAGTGTACTTCTC 1091
246 gCysSerArgProGluProPro.....ArgGlnHisPheGlyGlnArgI 261
1092 ATCAAGCTGAGCCAGAGAGTGAAGTCATTTGAAGAGAGAGTTGGAAAAAG 1141
261 lLeuValIleCysLeuSerLeuLeuLySLeuGlu..... 273
1142 TCCTGTCAAGTGCATGATTTATCTTTCATTTGCAATGCTGTGTCCT 1191
274IleGluProIlePheGlyIleLe 281
1192 GAAATGAAGAAGACCCACTACAAATGTGACCTTCTTGTACTCT 1241
281 uAlaLeuYrAspValArgGluLySlySlyIleSerGluAsnPheYrP 298
: : : : :
1242 ATCCCTGTTGACATAAATACAAACCGGAGATTTCTGCCGATTTCACAG 1291
298 heAspLeuAsnSerAspSerMetLyS..... 306
1292 TAGACCTGACCACTTCTCAGTGAAGCAAAATGATGCCACACGTCCTCC 1341
307 GlyLeuLeuArgAlaHisGly..... 313
1342 GCGGTGATGAATGGCAGTGGCGCGGAACCAATCTGCCCTCAGGGGAT 1391
314 .ThrHisProAlaIleSerThrLeuAlaArgSerAlaIlePheSerValT 330
1392 CCTTCATGAGCGCCATGACGATCCGAAACAGGAGGATATTTCAGTCA 1441
330 hTyrProSerProAspIlePheLeuValIleLySLeuGluLyValLeu 346
||| : : : : :
1442 CTGTCTCATTCAGATATATTCTTGTGGCCAGATTTGAAAAAGTCTT 1491
347 GlnGlnGlyAspIleSerGluCysCysGluProTyrMetValLeuLySgl 363
1492 ...CAGGGAGCATCACATTTGCGCTGAGCATATATG.....AAAG 1532
363 uValAspThrAlaLySAsnLySgluLySLeuGluLySLeuArgLeuAla 380
: : : : :
1533 TTCAGACTCTTCTAAGTGCGCCACAGAGTGTCTGAG.....AATG 1573
380 lAgLuglPheCysThrArgLeuGlyArgYrArgMetProPheAlaIaYr 396
1574 CCAGCAGGATGCCAAAGACTAGGACAGTATAGATGCCATTTGCTTGG 1623
397 ThrAlaValHisLeuAlaAsnIleValSerSerAlaGlyGlnLeuAspAr 413
1624 GCAGCAAGACATTG.....TTTAAGATGATGCTGGAATCTTGACAA 1667
413 gAspSerAspSerGluGlyGluArgArgProAlaTrpThrAspArgArg 430
1668 AAATGCC..... 1674
430 rArgArgLyProGlnAspArgAlaSerSerGlyAspAspAlaCysSerPhe 446
1675AGATTT 1680
447 SerGlyPheArgProAlaThrLeuThrValThrAsnPhePheLySglngl 463
1681 TCTGCG.....ACTACAGCAAGA 1700

```

463 uAlaGluArgLeuSerAspGluAspLeuPheLysPheLeuAlaAspMetA 480
      ::::::::::::::::::::|
1701 CAGCAATTAAGCTATCCATATGATGATGCTCAAGTACTGCAAGCTTTC 1750
      ::::::::::::::::::::|
480 rGArgProSerSerLeuLeuArgArgLeuArgProValThrAlaGlnLeu 496
      ::::::::::::::::::::|
1751 GGAACCTGGAAG...ATGGCTAAGCTCCCACTGATTTTAGCAATCTA 1797
      ::::::::::::::::::::|
497 LysLeuAspLeu...SerProAlaProGluAsnProHisPheCysLeuSe 512
      ::::::::::::::::::::|
1798 GACATTTCATTTGATATATGTTTCTCGAGACTTCCCTAATATATGTAA 1847
      ::::::::::::::::::::|
512 rProGluLeuLeuHisLysLysProTyrProAspProArgGlyArgPro 529
      ::::::::::::::::::::|
1848 ATCATACATT.....ATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1861
      ::::::::::::::::::::|
529 hrlYsgLile.....LeuGluPheProAlaArg 538
      ::::::::::::::::::::|
1862 CAACACATTTGAAACCTGCAGTAAACCTCCATCACGTTTGAGTGAG 1911
      ::::::::::::::::::::|
539 GluValTyr.....AlaProHisThrSerTyrAr 548
      ::::::::::::::::::::|
1912 GAATTGTGCGCTGCATACCAAAACACACCTTACACCATCTACAC 1961
      ::::::::::::::::::::|
548 gaSnLeuLeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln 564
      ::::::::::::::::::::|
1962 CAATCACCTTTACGTTTATCCCTAAGTACTGAAATAGACAGTCAAGAG 2011
      ::::::::::::::::::::|
565 ....GlySerValArgAsnLeuAlaValArgValGlnTyrMet...Thr 578
      ::::::::::::::::::::|
2012 CTTTTCGCAAGGCTAGAAATATTCGATTGCAATTCAATTCAAAGTTCA 2061
      ::::::::::::::::::::|
579 GlyGluAspProSerGlnAlaLeuProValIlePheGlyLysSerCys 595
      ::::::::::::::::::::|
2062 GATGAGGAGAACCTCAGCCCTTAAGTCATTATGCGACAGACCTGTGG 2111
      ::::::::::::::::::::|
595 sSerGluPheThrArgGluAlaPheThrProValIleTyrHisAsnLys 612
      ::::::::::::::::::::|
2112 GCCAGTTTTCACAAAGACGCTTGTGCTGAGTTTACACACACCAAA 2161
      ::::::::::::::::::::|
612 ePProGluPheTyrGluGluPheLysLeuHisLeuProAlaCysValThr 628
      ::::::::::::::::::::|
2162 ACCCAGATTTTATGATGAGATTAATAAGAGTGGCCCACTCAGCTGCAT 2211
      ::::::::::::::::::::|
629 GluAsnHisLysLeuLeuPheThrPheTyrHisValSerCys... 642
      ::::::::::::::::::::|
2212 GAAACACACACCTGTTGCTCACATTTCTCAGTCACTGAGACAATC 2261
      ::::::::::::::::::::|
643 .....GlnProArgProGlyThrAlaLeuGlnThrProValGlyP 656
      ::::::::::::::::::::|
2262 AACTAAAGCAGACGAAAGAGGATGCTGTTGAACCCAAAGTTGGCT 2311
      ::::::::::::::::::::|
656 heThrTyrIlePheLeuLeuGlnHisGlyArgLeuArgThrGlyProPhe 672
      ::::::::::::::::::::|
2312 ACTCCGTGCTCCCTCTGAAAGACGGAAGGTGTGCTCAAGCGCAG 2361
      ::::::::::::::::::::|
673 CysLeuProValSerValAspGlnProProProSerTyrSerValLeuTh 689
      ::::::::::::::::::::|
2362 CACATCCCGTCTCGGGAACCTT....CCTCGGCTATCTTGCTCA 2405
      ::::::::::::::::::::|
689 rProAspValAlaLeu.....ProGluMetArgTyrVal 701
      ::::::::::::::::::::|
2406 CCAAGGAGCTTGGATGGCAGGCAATATGCTCGGAATTAATTAAGGG 2455
      ::::::::::::::::::::|
701 spGlyHisLysGlyValPheSerValGluLeuThrAlaValSerSerVal 717
      ::::::::::::::::::::|
2456 ATGGAGCAGCAACGACGCGAAATTTCCATCATCTGTTTCTCAAGG 2505
      ::::::::::::::::::::|
718 HisProGluAspProTyrLeuAspLysPhePheThrLeuValHisValLe 734
      ::::::::::::::::::::|
2506 TATACTCAGGATCAGCATTTACATTAATTTTCCAGTACTGTCAGAAAA 2555
      ::::::::::::::::::::|
734 uGluGluGlyAlaPheProPheArgLeuLysAspThrValLeuSerGlu 751

```

```

2556 GCAATCTGAGACC.....CAAGCTTAG 2578
      ::::::::::::::::::::|
751 LysAsnValGluGlnLeuArgAlaSerLeuAlaAlaLeuArgLeuAla 767
      ::::::::::::::::::::|
2579 GAAC.....GACTTGAAGTACTTAAGAGCTCTCATGCGCATG 2619
      ::::::::::::::::::::|
768 SerProGluProLeuValAlaPheSerHisLysValLeuAspLysLeuVal 784
      ::::::::::::::::::::|
2620 GAAGCCACGATGATGCGCTTCTTGCCCACTATCCATTAACGAGCTT 2669
      ::::::::::::::::::::|
784 LArgLeuValIleArgProProIleLeuSerGlyGlnIleValAsnLeuG 801
      ::::::::::::::::::::|
2670 CCGAGTCTCACACAGA...GCCACACAGAGAAGATCGGTTAAAGTGA 2716
      ::::::::::::::::::::|
801 LArgGlyAlaPheGluAlaMetAlaHisValValSerLeuValHis... 816
      ::::::::::::::::::::|
2717 CTCGG.....GTCAATTATTCATGTTGTTGCCAGTGCATGAG 2754
      ::::::::::::::::::::|
817 ArgSerLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAl 833
      ::::::::::::::::::::|
2755 GAAGATTTGGAGAC.....CACTTGAG 2777
      ::::::::::::::::::::|
833 aAlaTyrValHisTyrAlaPheArgLeuProGlyThrGluProSerLeuP 850
      ::::::::::::::::::::|
2778 GTCAATGTGTTAAGTACGCGTAAAG.....GCTGAGCCATATGTT. 2817
      ::::::::::::::::::::|
850 roAspGlyAlaProProValThrValGlnAlaIleThrLeuAlaArgGly 866
      ::::::::::::::::::::|
2817 ..... 2817
      ::::::::::::::::::::|
867 SerGlyArgProAlaSerLeuTyrLeuAlaArgSerLysSerIleSer 883
      ::::::::::::::::::::|
2818 .....GCCTGGAATAC..... 2829
      ::::::::::::::::::::|
883 rSerAsnProAspLeuAlaValAlaProGlySerValAspAspGluVal 900
      ::::::::::::::::::::|
2829 ..... 2829
      ::::::::::::::::::::|
900 eArgGlyLeuAlaSerLysLeuHisGluGluLeu.....AlaLeu 914
      ::::::::::::::::::::|
2830 .....AAGACATGCATGAAGAACAGACCAATTCATG 2862
      ::::::::::::::::::::|
915 GlnTyrValValSerSerAla.....ValArgGluAlaIleLe 928
      ::::::::::::::::::::|
2863 ACCACGATTTCAAGCCCTTCGCGAATTTCTCTACCGCAACAAATCT 2912
      ::::::::::::::::::::|
928 uGlnHisAlaTyrPhePhePheGlnLeuMetValLysSerMetAlaLeu 945
      ::::::::::::::::::::|
2913 GAAGTACATGATGTTTCTTGTGATGATGATGATGATGATGATGATGAT 2962
      ::::::::::::::::::::|
945 LsLeuLeuLeuGlyGlnArgLeuAspThrProArgLysLeuArgPhePro 961
      ::::::::::::::::::::|
2963 ATTGATATGAGAACTCCAAAGTAAATGCTGCGCAACCAAGATTTCT 3012
      ::::::::::::::::::::|
962 GlyArgPheLeuAspAspIleThrAlaLeuValGlySerValGlyLeuG 978
      ::::::::::::::::::::|
3013 GCATCTATCATCATGCAAGTGAACCGTTGTAAATGCTGATGCCACA 3062
      ::::::::::::::::::::|
978 uValIleThrArgValHisLysAspValGluLeuAlaGlnHisLeuAsn 995
      ::::::::::::::::::::|
3063 CATCACTCAGAAAGTTTGAGATTAATCCAGAGGCACTTAACAGCGCAT 3112
      ::::::::::::::::::::|
995 LAsrLeuAlaPhePheLeuSerAspLeuSerLeuValAspArgGly 1011
      ::::::::::::::::::::|
3113 ATACCTTTGCTGCTTATCATCAAGATGTTTCACTTACATGACAGGGCC 3162
      ::::::::::::::::::::|
1012 PheValPheSerLeuValArgAlaHisTyrLysGlnValAlaThrArgLe 1028
      ::::::::::::::::::::|
3163 TTGCTCTTC.....AAGCAGATCAACAAATCAT 3191
      ::::::::::::::::::::|
1028 uGln.....SerSerProAsnProAlaAlaLeuLeuThrLeuArgMet 1043
      ::::::::::::::::::::|

```


5720 TCCGCCGCTTCATCTTTGAGATGCGCATTTTACGACAGCCGGAGAGCGCA 5769
1864 GlycylLeuProGluInHsLysArgLysThrLeuLeuSerThrAsp 1880
5770 GCGCGGGGTGGAAGAGCATGCAAAAGCGCAGCACCATCCTGCACGCCATACA 5819
1880 sAlaPheProTyrLeuLysThrArgIleArgValLysHsArgGluLnt 1897
5820 CTGCTTCCTTATGTGAAGAAGCCATCCCTGTCAATGTAACACACCACA 5869
1897 hrValLeuThrProValGluValAlaIleGluLysPheMetGlnLysThr 1913
5870 CTGACCTGAACCCCATGACGTGGCCATTGACAGATGAGTAGAAGGTG 5919
1914 ArgGluLeuAlaPheAlaThrGluGlnAspProProAspAlaLysMetIe 1930
5920 GCGGAGCTCCGGCAGCTGTGCTCTCGCGCGAGGTGACATGATCAACT 5969
1930 uGlnMetValLeuGlnGlySerValGlyProThrValAsnGlnGlyProL 1947
5970 GCAGCTCAAACTCCAGGCGAGCCTGACTGTTCAGTCAATGCTGCGCCAC 6019
1947 euGluValAlaGlnValPheLeuAlaGluIleProGluAspProLysIleu 1963
6020 TAGCATATGCGCGAGCTTCTTA.....GATGATCAAAACACA 6057
1964 PheArgHis...HisAsnLysLeuArgIleu.....CysPheLysAs 1976
6058 AAGCGATATCTGCACATTAAGTAGAAGAGCTTACGAAAGTTTTCAGGCA 6107
1976 pPheCysLysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleG 1993
6108 ATTGTGGAGAGCTTGCGGTCAACCTTACGGGTAAACGAAGCTCGATT 6157
1993 LyrProAspGlnLysGluTyrHisArgGluLeuGluArgAsnTyrCysArg 2009
6158 AAGAAGCCACGCTCGAGTATCAGGAAGAAATGAACCCAACTACAGGGA 6207
2010 LeuArgGluAlaLeuGlnProLeuLeuThrGlnArgIleProGluLeuMe 2026
6208 ATGGCGAAGAGCTTTCGAATCATGATCAGTAGACGATGTGCCCTCGA 6257
2026 tAlaProThrProProGlyLeuArgAsnSerLeuAsn 2038
6258 GGAGAAAGACGAGCGTC...TTACCGAATTCCTCTCAC 6291
seq_name: /SIDS1/gcdata/geneseq/geneseq-emb1/NA2000.DAT:AACT4524
seq_documentation_block:
ID AACT4524 standard; cDNA; 6816 BP.
XX
AC AACT4524;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORF79 polynucleotide sequence SEQ ID NO:157.
XX
XX Human; open reading frame; ORF; detection; cytostatic; hepatotropic;
XX vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antineoplastic;
XX antiviral; antibacterial; antifungal; antirheumatic; antihypoid;
XX antinaeemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antineoplastic; antineoplastic disease; coagulation;
XX thrombosis; contraceptive; ss.


```

3973 CAGCGTCCCTGCATCATCAACAACTGTGCACAGTGCACCGCTTAT 4022
1542 uArgAspSerThrPheAlaGluGlnValAlaAspLeuMetPheAsnLeuH 1559
4023 TAAGCAACCAAGCTTCTCTGATGTGAAGACATTAAACCAAAAGGATAC 4072
1559 lMetIleLeuThrAspThrValLysMetLysGlnHisGlnLysPro 1575
4073 GCACGCTGCTAATGGCCACCGCCAGATGAAGAGCATGAGACACGCCA 4122
1576 GluMetLeuIleAspLeuMetLysArgLLeuAlaArgGlyTyrGlnGlySe 1592
4123 GAGATGCTGGTGGACCTTCACAGCTGGCCAAATCTTATGGCACAC 4172
1592 rProAspLeuArgLeuThrTrpLeuGlnAspMetAlaGlyLysHisLag 1609
4173 GCCCGAGCTCAGGAAGACGTGCTGCACAGCATGGCCAGATCCATGTCA 4222
1609 luleuGlyAsnHisAlaGluAlaAlaGlnCysMetValHisAlaAlaAla 1625
4223 AAAATGGGATCTCTCAGAGGCACCAATGTCTATGTCCAGTAAACAGCC 4272
1626 leuValAlaGluTyrTrpLeuAlaLeuLeuGlnAspGlnArgHisLeuPro.. 1641
4273 CTAGTGGCAGATATCTCACAGGAAGAACGATCCAGTGGAGCCGCC 4322
1641 ..... 1641
4323 CCTTCTCCCCACAGCCATAGCGCTGCTGAGAGAGAGCCGGGAGCGC 4372
1642 ..... ValGlyCysValSerPheGlnAsnIleSerSerAsnValLeu 1655
4373 TGTTTAGCAAGAGATGCACCGCTTCAGGGCTATTACCCCAACATCGAC 4422
1656 GluGlnSerAlaIleSerAspAspIleLeuSerProAspGlnGluGlyPh 1672
4423 GAGGAGGCTCTCATGATGGAAGACGTGGGATGCAGGAT..... 4461
1672 ecysSerGlyLysHisPheThrGlnLeuGlyLeuValGlyLeuLeuGln 1689
4462 ..... GTCCATTTCAGAGAGATGTGCTGATGAGGCTCTTACG 4501
1689 lnaIlaAlaGlyTyrPheThrMetGlyLeuTyrGlnAlaValAsnGlu 1705
4502 AGTCGCGAGATGACCTCGAAGACCGGCTTACGAGCTCATTCGCGCAG 4551
1706 ValTyrLysAsnLeuIleProIleLeuGlnAlaHisArgAspTyrLys 1722
4552 ATCTACAAACTTATCATCCCATTTATGAGAGCGGAGGATTTTGAGAG 4601
1722 sleuAlaAlaValHisGlyLysLeuGlnGlnAlaPheThrLysIleMet 1739
4602 GCTGGCCCATCTGTATGACAGCGTGCACCGGCTTACAGCAAGTGACCG 4651
1739 lsgInserSerGlyTrpGluArgValAlaPheGlyThrTyrPheArgValGly 1755
4652 AGGTCATGCACTCGGCGCGAGGCTCTGGGGACTTCTCGGGTGGCC 4701
1756 PheTyrGly..... 1758
4702 TTCTTCGGGAGCGAGCGCATACCACTTTACAGACAGTGAACAGATGT 4751
1759 ... AlahisPheGlyAspLeuAspGlnGlnAlaPheValTyrLysGluP 1774
4752 GGAGGGATTTCTTGAAGATGAAGATGAAGAGATATATTACAGAGAAC 4801
1774 roserIleThrLysLeuAlaGluLysSerHisArgLeuGlnGluPheTyr 1790
4802 CCNAATCTACACCGCTGTGGAAATTTTCTCAGAGACTCTTAAGACTGTAC 4851
1791 ThrGluArgPheGlyAspAspValValGluIleLeuLysAspSerTyrPr 1807

```

```

4852 TCGGATAAATTGGTCTGAAAAATGTCAAAATGATACAGATTTGGCAA 4901
1807 oValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIleThrT 1824
4902 GGTCAACCTTAAGAGATCTGATTTCTAAGTATGCTTACATCCAGGTAC 4951
1824 yValGluProTyrPheAspThrTyrGlnLeuLysAspArgValThrTyr 1840
4952 ACSTCATCCCTCTTGTGAGAAAGAGTTGCAAGAGAAAGAAACAGAG 5001
1841 PheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProPheH 1857
5002 TTTGAGAGATCCCAACAATCCGCCCTTCATGATTTGAGATCCATTAC 5051
1857 rProAspGlyArgAlaHisGlyLeuLeuProGlnGlnHisLysArgLysT 1874
5052 GCAGACCGGGAAAGAGCAGGCGGGTGGAAAGACAGTCCAAACGGCGA 5101
1874 hIleLeuSerThrAspHisAlaPheProTyrIleLysThrArgIleArg 1890
5102 CCATCCTGACAGCATACACTGCTTCCCTTATGTGAAGAGCATCCCT 5151
1891 ValCysHisArgGlnGluThrValLeuThrProValGluValAlaIleG 1907
5152 GTCATGTACCAAGCACCACTGACCTGAACCCCATCGAGTGGCCATTGA 5201
1907 uAspMetGlnLysLysThrArgGluLeuValAlaPheAlaThrGlnLysAsp 1924
5202 CGAGATGATGAAGAAAGCTGGCGGAGCTCGGACAGCTGCTCTCGCGC 5251
1924 rObProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValGlyPro 1940
5252 AGGTGACATGATCAACAACCTGACAGCTCAACCTCAGGCGAGCTGAGT 5301
1941 ThrValAsnGlnLysProLeuGlnValAlaGlnValPheAlaGluIle 1957
5302 CAGGTAAATGCTGGCCACTGACATATGCGGAGCTTCTTA..... 5343
1957 eProGluAspProLysLeuPheArgHis... HisAsnLysLeuArgLeu 1972
5344 ... GATGATACAAACACAAAGCATATCTTGACATTAAGAGAGCTGC 5389
1973 ..... CysPheLysAspPheCysLysLysCysGlnLysPalaLeuArg 1986
5390 TTAAGGAAGTTTTCAGCAATTTGGGAAGCTTGGCGTCAACCTTACCG 5439
1987 LysAsnLysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGlu 2003
5440 GTRAAAGAACGCTGTGATTTAAGAGACCGCTCGAGTATCAGAGAAAT 5489
2003 uGluArgAsnTyrCysArgLeuArgGlnAlaLeuGlnProLeuLeuThrG 2020
5490 GAAAGCAACTACAGGAAATGCGAAGAGACTTCTGAAATCATGATG 5539
2020 lArgLeuProGlnLeuMetAlaProThrProProGlyLeuArgAsnSer 2036
5540 AGGAGATCTGCCCTGAGAGAAAGACGAGCTC... TTACCGAATTC 5586
2037 leuAsn 2038
5587 CTTCAC 5592

```

seq_name: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:AAH9572

seq_documentation_block:

ID AAH9572 standard; cdna; 1652 bp.

XX AAH9572;

XX AC 16-Oct-2001 (first entry)

XX DT

XX Human protein encoding cdna sequence seq ID NO:407.

XX

Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antirheumatic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; vitreous; anti-HIV; fungicide; antitumor; cardiovascular; antineoplastic; anemia; antiaggregant; haemostatic; vulnery; antileuk; osteoporosis; eczema; dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antineoplastic; rheumatoid arthritis; septic shock; pancreatitis; cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune; genetic disease; hematopoietic disorder; platelet disorder; asthma; thrombocytopenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder; ss.

Homo sapiens.
MO200153455-A2.
26-JUL-2001.
22-DEC-2000; 2000WC-US35017.
23-DEC-1999; 99US-0471275.
21-JAN-2000; 2000US-0488725.
25-APR-2000; 2000US-0552317.
(HYSE-) HYSEQ INC.
Tang YF, Liu C, Drmanac RT;
WPI: 2001-457603/43.
P-PSDB; AAM25631.
Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
Claim 1; Page 505; 1217p; English.

AAH99166 to AAH99994 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antirheumatic; immunosuppressive; antibacterial; endocrine; cardiac; central nervous system; vitreous; anti-HIV; fungicide; antitumor; cardiovascular; antineoplastic; antiaggregant; haemostatic; vulnery; antileuk; osteoporosis; eczema; dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis of disorders associated with the activity of a protein e.g. inflammation, rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction, neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal infections, autoimmunity, genetic diseases, haematopoietic disorders, anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers, osteoporosis, severe combined immunodeficiency, eczema, allergic rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression, Alzheimer's disease, Parkinson's disease, neurodegenerative and neurological disorders.

Sequence 1652 BP; 377 A; 478 C; 479 G; 318 T; 0 other;

alignment_scores:
Quality: 2195.00 Length: 427
Ratio: 5.177 Gaps: 2
Percent Similarity: 99.297 Percent Identity: 99.063

alignment_block:
US-09-736-968a-2 x AAH99572 ..

Align seg 1/1 to: AAH99572 from: 1 to: 1652

1623 AAlaAlaLeuValAlaGluTyrLeuAlaLeuGluLysGlnArgH1 1639
|||||
2 GCGGCGCGCTCGTGGCTGAGTACCTCGCCCTGCTGAGGACCCGCCA 51
1639 sLeuProValGlyCysValSerPheGlnAsnIleSerSerAsnValLeuG 1656
|||||
52 CCGGCCGCGGCTGCGCTTCCTTCAGAACATCTCATCAATGCTGCTAG 101
1656 lGluSerAlaIleSerAspAlleLeuSerProAspGluGluGlyPhe 1672
|||||
102 AGGAGTCCGCCATCTCCGACGACATCTCTGCGCCGACGAGAGGCTTC 151
1673 CysSerGlyLysHisPheThrGluLeuGlyLeuValGlyLeuGluGlu 1689
|||||
152 TGCTCCGGGAGACACTTCATGAGCTGGGCGCTTACGAGCGGTGATAGG 201
1689 nAlaAlaGlyTyrPheThrMetGlyGlyLeuTyrGluAlaValAsnGlu 1706
|||||
202 GCGAGCGCGCTACTTCACATGGGCGGCTCTTACGAGCGGTGATAGG 251
1706 aTyrLysAsnLeuIleProIleLeuGluAlaHisArgAspTyrLysLys 1722
|||||
252 TCTACAAACACCTCATCCCATCTCGAAGCCACCGCTGACTACAAAG 301
1723 LeuAlaAlaValHisGlyLysLeuGlnGluAlaPheThrLysIleMetH1 1739
|||||
302 CTGGCGCGGCTGCACGCGCAACTGCAGAGGCTTCCACCAAGATCATGCA 351
1739 sGlnSerSerGlyTyrProGluArgValPheGlyTyrThrPheArgValGlyP 1756
|||||
352 CGAGGTTCGCGCTGGAGCGGTGTGGAGCTATTTCCGCTGGGCT 401
1756 heTyrGlyAlaHisPheGlyAspLeuAspGluGlnGluPheValTyrLys 1772
|||||
402 TCTACGCGCGCCACTGCGTGCTGACCTGATGAGCAGAGATTGTGTACAG 451
1773 GluProSerIleThrLysLeuAlaGluIleSerHisArgLeuGluGluP 1789
|||||
452 GAGCATGATGATCACGAACTGGCAGAGATCTCACACCGCGCTGAGGAGTT 501
1789 eTyrThrGluArgPheGlyAspAspValAlaGluIleIleLysAspSerT 1806
|||||
502 CTACACGAGAGATTGGCGCAGCGCTGTTGAGATTAATCAAGACTCTA 551
1806 yProValAspLysSerLysLeuAspSerGlnLysAlaTyrIleGlnIle 1822
|||||
552 ACCGTGGACAAAGTCAAGCTTGACTCACAAAGGCTCATCATCAGATC 601
1823 ThrTyrValGluProTyrPheAspThrTyrGluLeuLysAspArgValTh 1839
|||||
602 ACGATGTGGAACCTTACTTTGATTACTTACAGCTCAAGACCGGGTAC 651
1839 rTyrPheAspArgAsnTyrGlyLeuArgThrPheLeuPheCysThrProP 1856
|||||
652 CTACTTTACCGCAACTATGSGCTTCCACATTTCTTCTTCCACGCGCT 701
1856 heThrProAspGlyArgAlaHisGlyLeuLeuProGluGlnHisLysArg 1872
|||||
702 TCACGCGCGATGGCGCGCACACGCGGAGCTGCCGAGCAACACACGCT 751
1873 LysThrIleLeuSerThrAspHisAlaPheProTyrIleLysThrArgI1 1889
|||||
752 AAGAGGCTGCTGACGACGACGACCTTCCCTTCAATCAAGATCCCAT 801
1889 eArgValCysHisArgGluGluThrValLeuThr.ProValGluValAla 1905
|||||
802 CCGTGTGTGCCACCGGAGGAGGAGGTGTGTACCCGACAGTGGAGTGGCC 851
1906 lIleGluSpMetGlnLysLysThrArgGluLeuAlaPheAlaThrGluG1 1922
|||||
852 ATCGAGGACATGCAAGAGACACGAGGAGCTGCGCTTTCGACACCGAGA 901


```
1922 nAspProProAspAlaLysMetLeuGlnMetValLeuGlnGlySerValG 1939
|||||
902 GGACCCGACGAGATGCTAGATGCTACAGATGCTGCTTCAGGGCTCTGTAG 951
|||||
1939 LyProThrValAsnGlnGlyProLeuGlnValAlaGlnValPheLeuAla 1955
|||||
952 GGCCCGACCGTGAACCAAGGGTCCCTGGAGGTGGCCAGGTGTTTTCAGCA 1001
|||||
1956 GluIleProGlnAspProLysLeuPheArgHisHisAsnLysLeuArgLe 1972
|||||
1002 GAGATCCCGGAGAGACCCCAAGCTCTTCCGGCATCACACAAATTGCGGCT 1051
|||||
1972 uCysPheLysAspPheCys.LysLysCysGlnAspAlaLeuArgLysAsn 1988
|||||
1052 CTGCTTCAGGACTTCTGCAAAGAAATGTGAGATGCGCTGCCGAAATAAT 1101
|||||
1989 LysAlaLeuIleGlyProAspGlnLysGluTyrHisArgGluLeuGluAr 2005
|||||
1102 AAGGCCCTGATTGGCCGGACCAAGAGAGTACCACCGTGAGCTGGAGCG 1151
|||||
2005 gAsnTyrCysArgLeuArgGluAlaLeuGlnProLeuLeuThrGlnArgL 2022
|||||
1152 CAACCTACTGCGCGCTGGGGAGGCTCTGCAGCCCTGCTTACCACAGCGCC 1201
|||||
2022 euProGlnLeuMetAlaProThrProProGlyLeuArgAsnSerLeuAsn 2038
|||||
1202 TGCCCCGAGCTGATGCACCCACCCGCGCTCAGAACTCCTTGAAC 1251
|||||
2039 ArgAlaSerPheArgLysAlaAspLeu 2047
|||||
1252 AGAGCAAGTTTCCGAAAGCGACACTC 1278
```

OM of: US-09-736-968a-2 to: Issued_Patents_NA:* out_format : pfs
Date: Oct 5, 2002 3:03 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frim+ .p2n.model -DEV=xlh
-O=/cgn2_1/USPFO.spool/US09736968/rnmat_02102002.145334.1346/app_query.fasta_1.2139
-DB=Issued_Patents_NA -OEM=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosome2
-TRANS=human40.cdi -LIST=45 -DOCCALIGN=200 -THR_SCORE=Pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORML=HEAPSIZE=500 -MIMLEN=0 -MAXLEN=200000000
-USER=US09736968 -CGN1_1.121 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-736-968a-2
Query length: 2047
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 113.010000

seq_list:
Seq: Stnd Orig ZScore EScore Len Documentation
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-258-261B-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-837-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-342-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-646A-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-458-076A-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-764-233A-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-457-335A-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-729-214-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-028-934-6 + 221.00 297.79 6.7e-09 28958
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-764-233A-1 + 221.00 289.69 1.9e-08 49377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-588-985-1 + 220.00 318.66 4.6e-10 6519
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-971-988-1 + 220.00 318.66 4.6e-10 6519
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-923-137-2 + 191.00 241.32 9.4e-06 36519
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-227C-7 + 182.00 222.48 0.0001 44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-804-198-1 + 182.00 222.48 0.0001 44377
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-147-236-1 + 177.00 228.37 4.9e-05 16836
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-147-236-10 + 177.00 228.37 4.9e-05 16836
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-144-085-3 + 173.00 210.85 0.0005 33529
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-443-501A-2 + 163.50 182.49 0.0177 71989
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-103-840A-1 + 162.50 118.23 67.28 4411529
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-07-689-008-1 + 160.50 207.88 0.0007 9540
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-428-517-1 + 158.50 178.92 0.0280 50937
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-19 + 154.50 176.11 0.0402 38506
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-330-409-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-102-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-969-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-480-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-486-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-472-1 + 154.50 167.30 0.1243 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-368-530-1 + 153.00 197.68 0.0025 7812
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-171-461-1 + 153.00 174.50 0.0726 43804
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-335-409-1 + 153.00 164.66 0.1746 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-102-1 + 153.00 164.66 0.1746 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-969-1 + 153.00 164.66 0.1746 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-480-1 + 153.00 164.66 0.1746 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-568-472-1 + 153.00 164.66 0.1746 68750
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-624-693A-16 + 152.50 212.45 0.0004 2787
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-09-320-878-21 + 152.50 200.88 0.0017 5970
/cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-125-468-1 + 152.50 176.37 0.0389 30001

/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-474-933-1 - 152.50 176.37 0.0389 30001
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-752-760A-1 - 152.50 173.99 0.0527 35081
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-443-501A-2 + 152.50 163.08 0.2139 71989
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-09-029-603-4 + 152.00 199.71 0.0019 6085
/cgn2_6/ptodata/2/ina/5B.COMB.seq:US-08-804-227C-1 + 152.00 169.92 0.0889 43280

seq_name: /cgn2_6/ptodata/2/ina/5A.COMB.seq:US-08-258-261B-6

seq_documentation_block:

Sequence 6, Application US/08258261B
Patent No. 5639949

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Unnes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

TITLE OF INVENTION: antipathogenic substances

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/258,261B

FILING DATE: 08-JUN-1994

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/457,205

FILING DATE: 01-JUN-1995

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

REFERENCE/DOCKET NUMBER: CGC 1506/CIP3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8614

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 28958 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

US-08-258-261B-6

alignment_scores:

Quality: 221.00 Length: 1480

Ratio: 0.350 Gaps: 78

Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:

US-09-736-968a-2 x US-08-258-261B-6 ..

Align seg 1/1 to: US-08-258-261B-6 from: 1 to: 28958

16 ThnValAlaAlaGluValAlaArgGlnValSerArgGluArgSerGlySe 32


```

551 LeuTyrValTyrTrpHisSerLeuAsnPheSerSerArgGln..GlySer 566
    :::: |||||::: |||||::: |||||::: |||||:::
11019 CCGAGAGCCCTGACCTCGCGGAGCAATCCCGCTCGGCGCGCGCTG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyLysProSe 583
    |||||::: |||||::: |||||::: |||||:::
11069 CCTCGCGCGAGCGCGGTGGCTTCTCT...TTACAGGCGCGCTCTCCCTC 11115
583 rGln.....AlaLeuPro..... 587
    |||||::: |||||::: |||||::: |||||:::
11116 GCAAGAGACCCCTGGCTCGAAGGSCATGCCGTCTCGGACACCATCTCT 11165
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
    |||||::: |||||::: |||||::: |||||:::
11166 ACCGGGACCGCGCTTCTCGAGCTCGCCCTCGACGTCGCCCGCGGTG 11215
602 AlaPheThrProValValTyrHisAsnLysSerProGluPheThrGlu 618
    |||||::: |||||::: |||||::: |||||:::
11216 GCCTCGACACCGCTC..... 11234
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
    :::: |||||::: |||||::: |||||::: |||||:::
11235 GCTACGCGCTCGAGGCGCCCTCGCTCGCTCCATCGACAGACCGCTCTC 11283
635 heThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
11284 ..... 11289
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnIleGlyArg.... 666
    |||||::: |||||::: |||||::: |||||:::
11290 ATCTCGCTCGGG.....CCCGTGGAGCGACGCGAGCAAGAGGC 11327
667 .....LeuArgThrGlyProPhe.. 672
    |||||::: |||||::: |||||::: |||||:::
11328 GCTCTCTTTCATAGCGGACAGAGAGAGCGGCTTACAGATGCGCCCTGA 11377
673 .....CysLeuProVal 676
11378 CTGCGCACGCGCAGCGGCTCTCTCGCGGCGGAGCCCATCCCTCTCGGC 11427
677 SerValAspGlnProProSerThrYrSerValLeuThrProAspValAl 693
    :::: |||||::: |||||::: |||||::: |||||:::
11428 GATCTCCACGACGAGCGCTCCCTCGAGTGCATC.....CCGGTGA 11468
693 aLeuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709
    |||||::: |||||::: |||||::: |||||:::
11469 CTGCGAGAGCGCTTACGCAACCTCGCCACACCTCGGCTTGCCTACGCGC 11518
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
    |||||::: |||||::: |||||::: |||||:::
11519 CCGAGTTCGAGGCGCTCGCTCGCTTACAAAGCGCGGCGAGAGCTCTTT 11568
726 LysPhePheThrLeuValHisValIleuGluGluValAlaPheProPheAr 742
    |||||::: |||||::: |||||::: |||||:::
11569 GCCGAGACCAAGCTCCCGGAAGCGCGCGAAAGAGATGCCGCCGGTTTGC 11618
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGluG 755
    |||||::: |||||::: |||||::: |||||:::
11619 CTTCCACCTCGCGCTCGCTCGAGCGCGCTGCACTGCGCTTGGAGG 11668
755 InGluLeuArgAlaSerLeuAla.....Ala 763
    |||||::: |||||::: |||||::: |||||:::
11669 ACGGAGGAGAGGAGCGTCTGCTGCTTCTCGCTTCTCGAGCGGAGTCTCG 11718
764 LeuArgLeuLysSerProGluProLeuValAlaPheSerHisValLe 780
    |||||::: |||||::: |||||::: |||||:::
11719 CTGCGCTCGCTCGCTCGACCATCTTGGCGGTGCTTCCAC..... 11760
780 uAspLysLeuValArgLeuValIleArgProProIleIleSerGlyGlnI 797
    |||||::: |||||::: |||||::: |||||:::
11761 .....GTTCCCAAGGGTGAATCTCGCTCT 11785

```

```

797 LeValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValValSer 813
    :::: |||||::: |||||::: |||||::: |||||:::
11786 CGATGCTCTGCGACGCGCGAGGTGACCTTGTCTCGGTGCAAGCG 11835
814 LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgG1 827
    |||||::: |||||::: |||||::: |||||:::
11836 CTGCGCATGCGGACGACGTCGCGCGCGAGCTCGGACCCCGGAGGCTCT 11885
827 yHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuProG 844
    |||||::: |||||::: |||||::: |||||:::
11886 CCACCATGATGCGCTC.....TTCCGCTCGACT 11914
844 LyThrGluProSerLeuProAspGlyAlaProProValThrValGluAla 860
    :::: |||||::: |||||::: |||||::: |||||:::
11915 GAGGAGAGCTCAAGCCCACT...TCACGCTTGGCGGCCCAAGCGGC 11961
861 AlaThrLeuAlaArgGlySer.....G1 868
    :::: |||||::: |||||::: |||||::: |||||:::
11962 GTCCTTCTCGGACAGGCGGCGACGATCTCGGCTCGACGCGCGCTGCG 12011
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSerS 884
    |||||::: |||||::: |||||::: |||||:::
12012 CCGCTACGCGCGACCTCGCTCGCTCGAAGCGCCCTCGACCGAGCGCTT 12061
884 eAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
    |||||::: |||||::: |||||::: |||||:::
12062 CGCTTCCGCGCTGCTGCTGCGCCCC...TTCACTGATCGACCGGCGGCG 12108
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
    :::: |||||::: |||||::: |||||::: |||||:::
12109 GACCTGCTCGCCGAGC.....GCCACGAG..... 12132
917 lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934
    :::: |||||::: |||||::: |||||::: |||||:::
12133 .....GCCACGCGCTCGACCTCGCTCGCTTGCAG...GCTGG... 12168
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln 950
    |||||::: |||||::: |||||::: |||||:::
12169 .....CTGCGCGAGCGAA 12180
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
    |||||::: |||||::: |||||::: |||||:::
12181 GCGCTCGCTCGCTCGCGC..... 12198
967 PileThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984
    |||||::: |||||::: |||||::: |||||:::
12199 .....CTGCTCTCTGTCACCCGACGCGCGCTCGCCACCGCACCGC 12238
984 lLysAspValGluLeuAlaGluHisLeuAsnAlaSerLeuAlaPhePhe 1000
    :::: |||||::: |||||::: |||||::: |||||:::
12239 AAGACGACGCTCAG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
    :::: |||||::: |||||::: |||||::: |||||:::
12253 .....GACCTGCG 12260
1017 lArgAlaHisTyrLysGluValAlaThrArgLeuGlnSer...SerProAs 1033
    :::: |||||::: |||||::: |||||::: |||||:::
12261 TCACGCGCGCTCTGCGGCGCTCGCGCTCGCGCAAGTGAAGCACCGAG 12310
1033 nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuGly 1050
    |||||::: |||||::: |||||::: |||||:::
12311 ACCTCCGCTCTT.....C 12324
1050 eHisGluHisTyrValThrLeuAsnLeuProCys..... 1061
    :::: |||||::: |||||::: |||||::: |||||:::
12325 CTGCTGACATGACCTCAGCGAGCGCTTCCAGCAGCGCTGCTAGGCGC 12374
1062 .....CysProLeuSerProPro.....AlaSerProSe 1071
    |||||::: |||||::: |||||::: |||||:::
12375 GCTGACACAGGAGAACGACGCTCGCTCGCAAGCGGAAACCCCTCA 12424
1071 rProSerValSerSerThrThr.SerGlnSerSerThrPheSerSerGln 1087

```

```

12425 TCCGAGGTTGGCGCAACGAGCTCGAGCGAGCGCTCATCCGCCCAA 12474
1088 A1aProAspProLySValThrSerMetPheGluLeuSerGlyProPheAr 1104
12475 GCACCCACGTCGGCCCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAlaLeuAlaL 1121
12493 .....CATATTCGACCAAGGACACCTTCGACGGCGCTCGCCCTCGTGC 12535
1121 euGluProGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
12536 ACCGCCCCCGAGGCCAG...GCGCCCTCGCACACGCGCAAGTCGCCATC 12582
1138 A1aValHisSerLeu..... 1142
12583 GCGGTGACGCGGAGGCTCACTTCGCGATGTGTCGACACCTTGG 12632
1143 .....LeuCysGlyHis..... 1146
12633 CATGTATCCGGCGAGCGCGCCGCTCGAGGCGCAAGCGCGGCATCG 12682
1147 ....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
12683 TTAAGTGAAGTCGGTCCAGGTGTCTCCGATACACGTAAGCGACGCGGTG 12732
1162 .....AlaGlu 1163
12733 ATGGGGGCTTCGGCGGAGCCTTGGTCCACGCGCATCGCGCCGCCCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTGCCCC.....ATCCCCACGCGCT 12808
1180 IsaAspPheAlaGluGly.ProGlyGlnArgSerArgLeuAla...SerLe 1195
12809 GGTCCTTCGCGCAAGCGCGGCTCCCATCATCATCTCATCCGCCCTAC 12858
1195 tLeuAspSerAspThrGluGlyLysAlaGlyThrIleAspN 1212
12859 TATGGACTCGTCGATCTCGGGC.....ATCTGAAGCCCATATCAGG 12899
1212 roserValAlaMet.....AlaIleAlaGlyLysProLeuAlaPro 1225
12900 TGTCTCTCATTCATCGCGCCCGCGCGCTCGGAGCGCGCTTCAGC 12949
1226 GLYSerArgLaserIleSerGlnGlyProProThrAlaSerArgAlaG 1242
12950 TCCGACGCCACCTCGCGCGCGAGGCTTTGGCCACGCCAGTCCAGGGAG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
13000 TGGAGCGCTTCGCGCGGCTCGGCTTCGAGCATGCGCCTCGGCTCTC 13049
1257 ..ValLeuTrpValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 ACCTGACCTGGGCTTCGAGCAGCATCTCTCGGCTCCACGCAATGGCGCG 13099
1271 ArgTrpAlaThrAspLeuThrLeu 1278
13100 GCATGATGTCTCTCGAGCTGTC 13123
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-456-837-6
seq_documentation_block:
; Sequence 6, Application US/08456837
; Patent No. 5643774
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven

```

```

; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip R.
; APPLICANT: Uknes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; NUMBER OF INVENTION: antipathogenic substances
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,837
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-456-837-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968A-2 x US-08-456-837-6 ..
Align seg 1/1 to: US-08-456-837-6 from: 1 to: 28958

16 ThrValAlaIaGluValArgLysGlnValSerArgLysSerGly 32
9487 ACACCTGGCATCGACGTC.....GTGTATACCCCGAGAGTTCGGCG 9527
32 rProHis.....SerSerArgArgCysSer...SerSerLeuGly 45
9528 CCGCCACGACAGATAGACATCCCGCGCGCGGCTCTCCGCTTCGGAC 9577
45 a1ProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
9578 TCTCGGACCAACAGCC.....CACGTATCTCTC 9606
62 SerArgProAspAlaGluProGlyProLeuArgAspLeuValGluPh 78
9607 GAGGAGGCTCCGCGCGCTGTGCGGC..... 9633

```

```

78 eProAlaAspSpleuGluLeuLeuLeuGluProAlaGluCysArgThr 95
9634 .....:|||||:
95 hrcGluProGlyIleProLys.....AspGluLysLeu 105
:|||||:
9656 GGTGGGAGCCGCTCCCGGGCGGTGCGCTGCTGCTGGCGGACGAGC 9705
:|||||:
106 AspAlaGluValArgAlaValAlaGluMetTyrIleGluAspTyrVal11 122
:|||||:
9706 GAGCGCGCGGTCCGCGCCAGGAGGCGG...CTCCGGACACCTCCT 9752
:|||||:
122 eValHisArgArgTyrGluTyrLeuSerAlaIleTyrSerProValThrT 139
:|||||:
9753 CGCCACAGAGACCTCGCCCTTATCGATGCGCTATCGACAGGCCACCA 9802
:|||||:
139 hrAspThrGln.....ArgGluArgGln 146
:|||||:
9803 CCGCGCGCCCTTCGAGACACCGCGCGCTCTCTGCGCGCGACCGGAC 9852
:|||||:
147 LysGlyLeu.ProArgGluVal..... 153
:|||||:
9853 GAGCTCCTCTCCGCGCTCGACTGCTGCGCCAGACAGCCGCCCGAG 9902
:|||||:
154 ....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
:|||||:
9903 CACCGTCTCGCGCGGAGCGGAGCCAGCAGCGATGCTTCTGCTTTC 9952
:|||||:
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThr 183
:|||||:
9953 CTGGGCAAGGCTCGAGGTGGAGGATGGCCCTCTCCCTGCTCGACTCC 10002
:|||||:
183 oArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAlaAla 200
:|||||:
10003 TCGCGCGGTCTCCGCGGTACG.....TCGAGCATGCGAGGCGC 10043
:|||||:
200 sPserLeuLeu.....ProSerLeuLeuGluArg 209
:|||||:
10044 GCTGCGCTCTCCAGTGAAGAGCTGCTGCGCTGCGCGCGGAGC 10093
:|||||:
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgArgGlnH 226
:|||||:
10094 AGGCGCGCCCTCC.....TCGACCGCGCTCGACGTCTGTA 10128
:|||||:
226 sArgProProAlaLeuLeuThrLeuTyrProAlaProAspGluAspGlu 242
:|||||:
10129 CAGCGCGCCCTCTTGGCGGTGATGCTCTCCCTGCGCGCTGCGGCG 10178
:|||||:
243 Ala.....ValGluArgCysSerArgProGluProProArgGluH 256
:|||||:
10179 GCTGCGGCTGACCGCGCGCGCTGCGCCACAGCCAGGCGGAGATCG 10228
:|||||:
256 sPheGlyGluArgIle.LeuValLysCysLeuSerLeuLysPheGluIle 272
:|||||:
10229 C.....CGCGCGCTTCTGCGAGGCGCTCTCTCTCGAGAGCGGCG 10272
:|||||:
273 GluIleGluProIle.....Phe 278
:|||||:
10273 CGCATGCGCGCTCGCGCAAGAGCGCTCACACCGTGGCGGCAACGG 10322
:|||||:
278 eGlyIleLeuAlaLeuTyrAspValArgGluLysLysIleSerGluA 295
:|||||:
10323 CGGCATGCGCGCGCTC..... 10338
:|||||:
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
:|||||:
10339 .....GAGCTGCGCGCTCCGACTCCAGACCTACCTACCTCGCTCC 10377
:|||||:
312 HisGlyThrHisProAlaIleSerThrLeuAla.....ArgSer 324
:|||||:
10378 TGGGGGAGCAGG.....CTCTCCACCGCGCGCTCACAGCCCGAGGGC 10421
:|||||:

```

```

324 rAlaIlePheSerValThrTyrProSerProAspIlePheLeuValIleL 341
:|||||:
10422 TACCTCGTATCCGCGAGCCCGCGCTGCGACGCTGCTC..... 10464
:|||||:
341 ySLeuGluLysValLeuGluGlnGlnLysAspIleSerGluCysGlyPro 357
:|||||:
10465 .....GAGCTCTCACCGCCGACCAAGGTG 10488
:|||||:
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
:|||||:
10489 TTCGCGCCAGATCCGGCTCGACTACGCTCCACATCCGCCAGATAGA 10538
:|||||:
374 uLysLeuArg.....LeuAlaIleGluGlnPheCysThrArgLeuGlyA 389
:|||||:
10539 GCGCTTCAGACGAGCTCGCGCA..... 10563
:|||||:
389 rGTYrArgMetProPheAlaThrPThrAlaValHisLeuAlaAsnIleVal 405
:|||||:
10564 .....GGCTTAGCCACATCGCT 10581
:|||||:
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGlyArgArg 422
:|||||:
10582 CCTCGAGGTGGAGCTCCCTTATTCGACCGTCACCGGACCAAGCT 10631
:|||||:
422 gProAlaThrPheAspArgArgArgGlyProGluAspArgAlaSerS 439
:|||||:
10632 CGAC.....GGCTTCG 10642
:|||||:
439 eGlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
:|||||:
10643 AGCTGAGCGCGCGTACTGCTATCGAAACTCCGGAACAGCTCTCTG... 10689
:|||||:
456 ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLe 472
:|||||:
10690 .....TTCTCGAGCGCGACGAGCGGCTCTCGAC...GATGG 10724
:|||||:
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
:|||||:
10725 GCATGCTTCTCCGAGGTGAGTCAAGCCCATCCGCTCACTGCTGCGCC 10774
:|||||:
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
:|||||:
10775 TCCGCGAGACCTCGCGAGGCTCACCGCTCGATCCGCTGCTGGGCTCC 10824
:|||||:
504 oGluAsnProHisPheCysLeuSerPro.....Glu.LeuLeuHisIle 518
:|||||:
10825 ATTGACGAGAGAGAGGACCATCCGCGCTGCTCTCTCTGCGGCGA 10874
:|||||:
519 LysProTyrProAspProArgGlyArg..... 527
:|||||:
10875 GCTCTCTACCGAGGCTCGCGCTGAGTGAGAGACTTCTTCGCGCCT 10924
:|||||:
528 .....ProThrLysGluIleLeuGluPheProAlaA 538
:|||||:
10925 AGCTCCCCCGAAGTCTCCCTCCAC.....CTACCCCTTCACAGGA 10968
:|||||:
538 rGluValTyrAlaProHisThrSerTyrArgAsnLeu..... 550
:|||||:
10969 GAGCGGTTCTGCTCGAGCTCTCCAGAGAGACGCTCCGACGTGCGCT 11018
:|||||:
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln...GlySer 566
:|||||:
11019 CCGAGGCTGCTCGCGCGACCAATCCGCTGCTGCGCGCGCGCTCG 11068
:|||||:
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProse 583
:|||||:
11069 CCTTGGCGGACCGCGGTGCTTCTCT...TTACAGGCGGCTCTCCTCC 11115
:|||||:
583 rGln.....AlaLeuPro..... 587
:|||||:
11116 GCGAGGACCCGTGCTCGAAGGCGCATCCGCTTTCGCGACACCAATCT 11165
:|||||:
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
:|||||:

```



```

1138 AlaValHisSerLeu..... 1142
1143 .....LeuCysAlaHis..... 1146
12583 GCGGTGCACGCGGACAGGCTCAACTTCGCGATGTCGACACCCCTTGG 12632
1143 ..... 1146
12633 CATGATATCCGGGACGACGCCGCCGCTCGAGGCGAAGCGCGGCGCATCG 12662
1147 .....AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
12683 TTTCTGAATCGGTTCAGAGTGTCTCCCGATACACCGTAGGCGACCGGGTGG 12732
1162 ..... 1163
12733 ATGGGGGTCTTCGGCGCACGCTTTGGTCCACGCGCATCGCGCGCGCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProAlaGlu 1180
12783 CATGATCTGCCCC.....ATCCCCACGCGCT 12808
1180 IsasppheAlaGluJly..ProGlyInArgSerArgLeuAla...SerMe 1195
12809 GGTCTTCCGCCCAAGCGCGCAGGCTCCCATCATCTCATACCGGCTTAC 12858
1195 IleuAspSerAspThrGluGlyGluGlyAspIleAlaGlyThrIleAsp 1212
12859 TATGACACTGTCGATCTCGGCG.....ATCTGAACCCACATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyGlyProleuAlaPro 1225
12900 TGTCTCATCATGCGCGCGCGCGCGCGCTCGGACGCGCGCTTCACG 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaG 1242
12950 TCCGACGCCACCTCGCGCGCGAGGCTTTCGCACGCCGACGTCAGGAG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuAlaCys..... 1256
13000 TGGAGGCGCTTCGCGCGCTCGGCTTCGACGATCGCGCCTCGCTC 13049
1257 ..ValleuTrpValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 ACGTGACCTGGGCTTCGAGCAGACACTTCCCGCGCTCCACGATGGCGG 13099
1271 ArgTrpAlaThrAspLeuThrLeu 1278
13100 GCATGATGTCGTCTCGACTGTC 13123
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-457-342-6
seq_documentation_block:
; Sequence 6, Application US/08457342
; Patent No. 5662898
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Unkes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA

```

```

; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,342
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1506/CIP3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ANTI-SENSE: NO
; US-08-457-342-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-08-457-342-6 ..
Align seg 1/1 to: US-08-457-342-6 from: 1 to: 28958

16 ThrValAlaAlaGluValArgLysGlnValSerArgLysArgGlySer 32
9487 ACACCTGCCATCGACGCTC.....GTGATACCCCGAGGTCTTGGCG 9527
32 ProHis.....SerSerArgArgCysSer...SerSerLeuGly 45
9528 CCGCCACGAGATGACGAGTCCCGCGCGCGCGCTTCGCGCTTCGAGC 9577
45 aProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
9578 TCTCCGCGACCAACGCC.....CACGTATCTCTC 9606
62 SerArgProProAspAlaGluProGlyProLeuArgAspLeuValGluPh 78
9607 GAGGAGGCTCCCGCGCGCTTCGCGG..... 9633
78 eProAlaAspAspLeuGluLeuLeuGlnProArgGluCysArgThr 95
9634 .....GAGCCCGCCACCTCATACAGCG 9655
95 hngLuproGlyLeuProLys.....AspGluLysLeu 105
9656 CCGCCGACCGCTCCCGCGCGGTGCGCTCTGCTCTGCGCGACGAGC 9705
106 AspAlaGluValArgAlaAlaValGluMetTyrIleGluAspTrpValI 122
9706 GAGGCGCGCTCCCGCGCGCGCGAGCGAAGCG...CTCCGCGACACCTCT 9752
122 eValHisArgArgTyrGlnTyrLeuSerAlaAlaTyrSerProValThr 139

```

```

9753 CGCCGAGACGACCTGCGCCCTTATGATGGGCTATTGCAAGGCCACCA 9802
139 hrAspThrGln.....ArgLuarGln 146
9803 CCGGCGCCCACTTCGACACCGCGCGCTCTCTGGCCGCGACCGCGAC 9852
147 LysGlyLeu, ProArgLysVal..... 153
9853 GACCTCTCTCCGCGCTCGCATGCTCGCCACGACACCGCCCGGAC 9902
154 ....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
9903 CACCGTTCTGGCGCGAGCGAGCCAGCGAGCTGTTCTGCTTTC 9952
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
9953 CTGGGCAAGGCTCGCAGTGGGAAGGATGGCCCTCTCCCTGCTCACTCC 10002
183 OAArgSerSerGlyAlaSerSerThrPheAspLeuArgAsnLysAlaAla 200
10003 TCCCGCGCTTTCGCGCTCAGC.....TCGAAGCATGCGAGCGGC 10043
200 sPserLeuLeu.....ProSerLeuGluArg 209
10044 GCTCGCTCTCAGCTGAGTGAAGCTGCTGCGCTGCGCGCGACG 10093
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgArgLys 226
10094 AGGCGCGCCCGCTCC.....TCGACCGCGCTCAGCGTCGTA 10128
226 sArgProProAlaLeuLeuThrLeuThrProAlaProAspGluAspGlu 242
10129 CAGCCCGCGCTTTCGCGCTGATGGTCTCTCGCGCGCTGCGCTC 10178
243 Ala.....ValGluArgCysSerArgProGluProArgGln 256
10179 GCTCGCGCTGAGCCCGCGCGCTGCGCGCACAGCGCGAGATGG 10228
256 sPheGlyGlnArgGlyLeuValLysCysLeuSerLeuLysPheGluLe 272
10229 C.....CGCGCGCTTCTGCGCAGGCGCTCTCCCTCGAGGACCGCGG 10272
273 GluLeuLeuProLe.....Ph 278
10273 CGCATCGCGCGCTGCGCGAAGCGCTCACACCGTCGCGCGACAGG 10322
278 eGlyLeuLeuAlaLeuThrAspValArgLysLysLysSerGlu 295
10323 CGGCAAGCGCGCGCTC..... 10338
295 snPheThrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
10339 .....GAGCTCGCGCGCTCGACCTCGACACCTACCTCGCTCC 10377
312 HisGlyThrHisProAlaLeuSerThrLeuAla.....Argse 324
10378 TGGGCGGACAG.....CTTCACCGCGCGCGCTCGACAGCGCGAGG 10421
324 ValAlaPheSerValThrThrProSerProAspThrPheLeuValIleL 341
10422 TACCTCGATCGCGGAGCGCGCGCGCTGCTC..... 10464
341 yLsLeuGluLysValLeuGlnGlnGlyAspThrLysCysGluPro 357
10465 .....GACGTCTCTACCGCGCACCAAGGTG 10488
356 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
10489 TTCGCGCCCAAGATCCGCGTCACTACGCCCTCCGCGCGAGATGA 10538
374 uLysLeuArg.....LeuAlaAlaGlnGlnPheCysThrArgLeuGly 389

```

```

10539 CGCCGTCACAGACGAGTCCGCCA..... 10563
389 rGlyArgMetProPheAlaThrPheAlaValHisLeuAlaAsnIleVal 405
10564 .....GGTCTAGCCAACTGCT 10581
406 SerSerAlaGlnGlnLeuAspArgAspSerAspSerGluGluArg 422
10582 CCTCGAGCTGCGAGAGCTCCCTTATTTCAGACCTCACCGGACACAGCT 10631
422 gProAlaThrPheAspArgArgArgGlyProGluAspArgAlaSer 439
10632 CGAC.....GGCTCCG 10642
439 eArgLysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
10643 AGCTCGAGCGCGCTACTGGTATGAAACCTCCGGCAACCGTCTC... 10689
456 ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLe 472
10690 .....TTCGAGCGCGACCGAGCGGCTCTCGAC...GATGG 10724
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
10725 GCATCGCTTCTCGTGAAGTCAAGCCCATCCCTGCTCAGCTCGCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysLysLeuSerProAlaPr 504
10775 TCGGCGAGACCTGCGAGCGCTCACCGCTCGATCCGCTGCTGCTCC 10824
504 oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
10825 ATTGAGAGAGAAGAACCACTCGCGCGCTGCTCTCTCGGGCGGA 10874
519 LysProGlyProAspProArgGlyArg..... 527
10875 GCTCTTACCGAGCGCTGCGCTGAGTGAAGGACTTTCGCGGCT 10924
528 .....ProThrLysGluIleLeuGluPheProAla 538
10925 ACGTCCCGCGAGAGTCTCCCTCCAC.....CTACCCCTTCAGCA 10968
538 rGluValTyrAlaProHisThrSerTyrArgAsnLeu..... 550
10969 GAGCGGTTCTGGCTGAGAGCTGTCAGAGCAAGAGTTCGAGTGGCT 11018
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln...GlySer 566
11019 CCGAGGCTGAGCTGCGCGACCAATCCGCTGCTGAGCGCGGCTCG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProse 583
11069 CTTTCGCGCGAGCGCGGTGGTTTCT...TTACAGGCGCGCTCTCC 11115
583 rGln.....AlaLeuPro..... 587
11116 GCAAGAGACCGCTGGCTGAAAGCCATGCGCTTGGGACACCATCT 11165
588 .....ValIlePheGlyLysSerSerGlySerGluPheThrArgGln 601
11166 ACCGCGACCGGCTTCTGAGCTGCGCGCTGACAGTCCGCGCGAGCT 11215
602 AlaPheThrProValValTyrHisAsnLysSerProGluPheThrGluG 618
11216 GCTTCGACACCTC.....GAAAG 11234
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLysLeuLeu 635
11235 GCTACAGCTGAGAGCGGCTGCTGCTCCATGCGAGAACGCGTCTC... 11283
635 heThrPheThrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
11284 .....CTCGAG 11289

```



```
12733 ATGGGGCTCTGGCGCAGCCTTTGGTCCAGCGCATCGCGGAGCCCG 12782
1163 uleuYrleuProLeuLeuSerIleAlaArgAspThrleuProArgleuH 1180
12783 CATGACTCTGCCCC.....ATCCCCCAGCGCT 12808
1180 isasphelaIagluYrProglYnArgSerArgleuAla...Serle 1195
12809 GGTCTTCCTCCAGCGCCAGCGTCCCATCATCATTCACCCGCTTC 12858
1195 cleuasSerAspThrIgluYrIgluYrAspIleAlaIgluYrleuAsp 1212
12859 TATGAGACTGTGATCTCGGCG.....ATCTGAACCCCAATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaIgluYrProleuAlaPro 1225
12900 TGTCTTCATCCAGCGCGCGCGCGCTCGGAGCGCGCGCTTCAGC 12949
1226 GlySerArgAlaSerIleSerIgluYrProThrAlaSerArgAlaI 1242
12950 TCCACAGCCACCTCGCGCGGAGGTTCCTCCACCGCCAGTCCAGGAG 12999
1242 ycysAlaLeuSerAlaIgluSerArgThrleuLeuAlaCys..... 1256
13000 TGGAGGCTCTCCGCGCTCGGCTTCGACGATGCGCAGCTCGGCTCTC 13049
1257 ..ValleuTrpValLeuYrAsnThrIglu.....ProAlaLeuIleu 1270
13050 ACCTGACCTGGGCTTCGAGCAGCATCTTCGCTCAGCATGCGCGCG 13099
1271 ArgTrpAlaThrAspLeuThrleu 1278
13100 GCATGATGTCTGCTCGACTGTC 13123

seq_name: /cgn2_6/ptodata/2/ina/5A-COMB.seq:US-08-457-646A-6

seq_documentation_block:
; Sequence 6, Application US/08457646A
; Patent No. 5679560
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James M.
; APPLICANT: Beck, James Joseph
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Gaffney, Thomas Deane
; APPLICANT: Lam, Stephen Ting
; APPLICANT: Hammer, Phillip E.
; APPLICANT: Unnes, Scott Joseph
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/457,646A
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION NUMBER: US 08/457,205
; FILING DATE: 01-JUN-1995
; APPLICATION NUMBER: 08/258,261
```

```
; FILING DATE: 08-Jun-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8689
; TELEFAX: 919-541-8684
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-08-457-646A-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968A-2 x US-08-457-646A-6 ..
Align seg 1/1 to: US-08-457-646A-6 from: 1 to: 28958

16 ThrValAlaAlaIgluValArgIgluValSerArgIgluArgSerGly 32
|||||
9487 ACACGCGCATGACGATC.....GTGATACCCCGAGGCTTGTGGC
32 rProHis.....SerSerArgGlySer...SerSerleuGly 45
|||||
9528 CGCCACAGATAGCATGCTCCCGCGCGCGGCTCTCCGCTTCGGAC
45 alProleuThrIgluValIgluProleuAspPheIgluAlaLeu 61
|||||
9578 TCTCCGCGCACCAAGCC.....CACGTATCTCTC
62 SerArgProProAspAlaIgluProGlyProleuArgAspLeuValIglu 78
|||||
9607 GAGGAGCTCCCGCGCGCTGCGGC.....
78 eProAlaAspAspLeuIgluLeuIleuGlnProArgIgluGlyAspThr 95
|||||
9634 .....GAGCCCGCACCTCACAGACG 9655
95 hrIgluProGlyIleProlys.....AspGluIglu 105
|||||
9656 GGTCCGACGCTCCCGCGCGGTGTGCGGCTCTGTCGCGCAGAC 9705
106 AspAlaIgluValArgAlaAlaValIgluMetYrIleIgluAspTrpVal 122
|||||
9706 GAGCCCGCTCCCGCGCGCGCGGCGGAGCG...CTCGGACACCTCTCT 9752
122 eValHisArgArgYrIgluYrleuSerAlaAlaIgluSerProValThr 139
|||||
9753 CGCCACGACGACGCTCGGCTTATGATGATGCTATTCGAGCGACCA 9802
139 hrAspThrGln.....ArgIgluArgGln 146
|||||
9803 CCGCGCGCGCTTCGAGCAGCGCGCGCTCTGCGCGCGGAC 9852
147 LysGlyLeuProArgIgluVal..... 153
|||||
9853 GAGCTCTCTCGCGCTGACTGCTGCGCCAGGACAGACCGCGCGAG 9902
154 .....PheIgluIgluAspAlaSerGlyAspIgluArgSerGlyProGlu 168
|||||
9903 CACGTTCTCGGCGCGGAGCGGAGGACGACGAGTGTCTTCTCTTTC 9952
```



```
11428 GATCTCCAGAGTGGCTCCCTCCAGATGCCATC.....CGGGGA 11468
      ::::: ||||| :::::
693 AleuProgluMetArgTrpValAspGluHisLysGluVal...PheSerV 709
      ||| ||||| ::::: |||||
11469 CCTCGAAGGGCTTACGGCAACCTCCGCAACCTCGGGCTTCCATACGGCC 11518
      ||||| ::::: |||||
709 ALGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
      ||||| ::::: |||||
11519 CCGAGTTCCAGAGGCTCCGCTCCCTCTACAAAGCGCGGCGAGAGACTCTT 11568
      ||||| ::::: |||||
726 LysPhePheThrLeuValHisValLeuGluGluValAlaPheProPheAr 742
      ||| ::::: |||||
11569 GCGCAAGCCAAAGCTCCCGAAGCGCGCAAAAGATGCGCGCGGTTCG 11618
      ||| ::::: |||||
742 GluLys.....AspThrValLeuSerGluGluValAsnValGluG 755
      ||| ||||| ::::: |||||
11619 CCTCCACCTCGCTCGTCCAGACGGCTCGATGCACTGGCTTTGAGG 11668
      ||||| ::::: |||||
755 InGluLeuArgAlaSerLeuAla.....Ala 763
      ||||| ||||| ::::: |||||
11669 ACGAGCAAGAGAGGAGCGGTCTGCCCTTCTGAGAGCGAGTCTCG 11718
      ||||| ::::: |||||
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
      ||||| ::::: |||||
11719 CTGGCTCTCGTGGTCCACCACTTGGCGCTTCCAC..... 11760
      ||||| ::::: |||||
780 uAspLysLeuValArgLeuValIleArgProIleIleSerGluIle 797
      ||||| |||||
11761 .....CGTCCCAAGGGTGAATCTCCGCT 11785
      ||||| ::::: |||||
797 IValAsnLeuGluArgGluValaPheGluAlaMetAlaHisValSer 813
      ::| ||||| ::::: |||||
11786 CGATGCTCTGGCGAGCGCGAGGTGACCTTGGCTGGTCCAGAGCG 11835
      ||||| ::::: |||||
814 LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgG 827
      ||||| ::::: |||||
11836 CTGGCGATGCGAGAGAGCTCGCGCGGACGTCGCAACCCGCGAGTTC 11885
      ||||| ::::: |||||
827 yHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuPro 844
      ||||| |||||
11886 CCACCATATGCGCTC.....TTCCGCTCGACT 11914
      ||||| ::::: |||||
844 IyThrGluProSerLeuProAspGluAlaProProValThrValGlnAla 860
      ::||| ::::: |||||
11915 GGAGCGAGCTCCAAAGCCCACT...TCACGCTCGCGCCGAGGAGGC 11961
      ||||| ::::: |||||
861 AlaThrLeuAlaArgGlySer.....G 868
      ::||| ::::: |||||
11962 GTCTTCTCGACAGCGCGGACGATCTCGGCTCGAGCGCCGCTGC 12011
      ||||| ::::: |||||
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerLeuSerS 884
      ||||| |||||
12012 CCGCTACGCCGACCTCGCTGCCCTCCGAAGCGCCCTCGACCAAGGCGCTT 12061
      ||||| ::::: |||||
884 eArgnProAspLeuAlaValaLarProGlySerValAspAspGluValSer 900
      ||||| |||||
12062 CGCCTCCGCGGCTCGTCTCGCCCC...TTCATGATCGACCGGCAAGGC 12108
      ||||| ::::: |||||
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
      ::||| |||||
12109 GACCTCGTCCGAGC.....GCCACAG..... 12132
      ||||| ::::: |||||
917 IValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPheP 934
      ::||| |||||
12133 .....GCCACCGGCTCGCATCTCGCCCTTGCDA...GCTGG.... 12168
      ||||| ::::: |||||
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGluGln 950
      |||||
12169 .....CTCGCGGAGGA 12180
      |||||
951 ArgLeuAspThrProAlaGlyLysLeuArgPheProGluArgPheLeuAspAs 967
      ||||| ::::: |||||
```

```
1212 roSerVala1aMet.....Ala1leAlaGlyGlyProLeu1a1aPro 1225
      ||| :||| ||| :||| |||
12900 TGTCTCATTCATGCGCGCCGCCGCGCGCGCGCGCGCGCGCGCGTTCAGC 12949
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1226 GlySerArgAlaSer1leSerGlnGlyProProThrAlaSerArgAlaG1 1242
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
12950 TCGCACGCCACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 12999
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13000 TCGAGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCTC 13049
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1257 ..ValleuTriValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
      ||| ||| :||| :||| :||| :||| :||| :||| :||| :|||
13050 ACGTGACCGCGCTTCGACGACGACTCTCGCGCTCCGCGTCCAGCATGGCGCG 13099
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1271 ArgTrpAlaThrAspLeuThrLeu 1278
      ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
13100 GCATGATGTCTCTCCGACTGTC 13123
```

seq_name: /cgn2_6/ptodata-a/2/lna/5A_COMB.seq:us-08-458-076a-6

seq_documentation_block:

Sequence 6, Application US/08458076a

Patent No. 5698425

GENERAL INFORMATION:

APPLICANT: Schupp, Thomas

APPLICANT: Ligon, James M.

APPLICANT: Beck, James Joseph

APPLICANT: Hill, Dwight Steven

APPLICANT: Ryals, John Andrew

APPLICANT: Gaffney, Thomas Deane

APPLICANT: Lam, Stephen Ting

APPLICANT: Hammer, Phillip E.

APPLICANT: Uknes, Scott Joseph

TITLE OF INVENTION: Genes for the synthesis of

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

NUMBER OF SEQUENCES: 22

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-458-076a-6

alignment_scores:
Quality: 221.00 Length: 1480
Ratio: 0.350 Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-08-458-076a-6

Align seg 1/1 to: US-08-458-076a-6 from: 1 to: 28958

```
16 ThrValAlaAlaGluValArgLysGlnValSerArgGluArgSerGlyse 32
      |||:||||| :||| ||| :||| ||| :||| ||| :||| ||| :|||
9487 ACACCTCCGACATCGACGTC.....GTGATATACCCCGAGGCTTGGGC 9527
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
45 alProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9528 CCGCCACGAGATAGACATCCCGCCGCGCGCGCGCGCGCGCGCGCGTTCGAGC 9577
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9578 TCTCCGCGACCCACGCG.....CACGTCATCTTC 9606
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
62 SerArgProProAlaGluProGlyProLeuArgAspLeuValGluPh 78
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9607 GAGAGAGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 9633
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
78 eProAlaAspAspLeuGluLeuLeuLeuGlnProArgGluCysArgThr 95
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9634 .....GAGCCGCGACCTCACAGACGG 9655
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
95 hrgLupProGlyIleProLys.....AspGluLysLeu 105
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9656 CGTCGCGACCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 9705
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
106 AspAlaGluValArgAlaAlaValAlaGluMetYrIleGluAspTrpVal 122
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9706 GAGGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 9752
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
122 eValHisArgArgTrgIntYrLeuSerAlaAlaYrSerProValThr 139
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9753 CGCCACGACGACCTCGCCCTTATGATGCGCTATTCGACGCGCACCA 9802
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
139 hraspThrGln.....ArgGluArgGln 146
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9803 CCGCGCGCGACCTTCGAGCACCGCGCGCGCGCTCTCGCGCGCGAGCGAGC 9852
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
147 LysGluLeu..ProArgGlnVal..... 153
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9853 GAGCTCTCTCCGCGCGCTCGACTCGCCACGACAGACGCCGCCGCGAGC 9902
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
154 .....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9903 CACGCTTCGCGCGCGAGCGAGACCAAGCGAGCGAGTGTCTGCTCTTTC 9952
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
9953 CTGGCAGAGCTCCAGTGGAGGATGGCGCTCTCCGCTCGACTCC 10002
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
133 GArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAlaAla 200
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
10003 TCGCGCGCTCTCCGCGCTCAGC.....TCGACGATCGGAGCGCGC 10043
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
200 spSerLeuLeu.....ProSerLeuLeuGluArg 209
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
10044 GTCGCTCTCAGCGTGAAGTGAAGCTGCTCGCGCTCGCTCGCGCGCGAGC 10093
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgGlnIh 226
```



```

10094 AGGCGCCCTCC.....TCGACCGCGTCGACGTCGTA 10128
      |||||
226 sArgProAlaLeuLeuThrLeuTyrrProAlaProAspGluAspGlu 242
      : |||||: ||| ||| |||
10129 CAGCCCGCTCTTGGCGTATGTCCTCCGCGCCCTCGCGCGCTC 10178
243 Ala.....ValGluArgCysSerArgProGluProProArgGluH 256
      ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10179 GCTCGCGCGTGAAGCGCGCGCTCGCGCGCACAGCGGAGATCG 10228
256 sPheGluArgGluLeuValLysCysLeuSerLeuLysPheGluIle 272
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10229 C.....CGCGCGCTTCGCGAGCGCTCTCTCCGAGAGCGCGCG 10272
273 GluLeuGluProIle.....Ph 278
      : |||
10273 CGCATCGCCCGCTCGCGAGAAAGCGTACCACCGTCGCGCGACGG 10322
278 eGlyIleLeuAlaLeuTyrrAspValArgGluLysLysIleSerGlu 295
      |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10323 CGCGATGCGCGCGCTC..... 10338
295 snPheTyrrPheAspLeuAsnSerAspSerMetLysGluLeuLeuArg 311
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10339 .....GAGCTCGCGCGCTCGACCTCGACCTACCTCGCTCC 10377
312 HisGlyThrHisProAlaIleSerThrLeuAla.....Argse 324
      ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10378 TGGGCGGACAGG.....CTCTCCACCGCGCGCTCACAGCGCGCGG 10421
324 rAlaIlePheSerValThrTyrrProSerProAspIlePheLeuValIle 341
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10422 TACCTCTGTAATCCGCGAGCGCGCGCTCGAGCGCGCTCTC..... 10464
341 yIleuGluLysValLeuGluGlnGlyAspIleSerGluCysGluPro 357
      |||||
10465 .....GAGCTCTCACCGCGCCACCAAGGTCG 10488
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeu 374
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10489 TTGCGCGGACAGATCCGCGTGCAGCTACGCTCCACTCGCGCGAGTGA 10538
374 uLysLeuArg.....LeuAlaAlaGluGlnPheCysThrArgLeuGlu 389
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10539 CGCGCTCCAGAGACGAGCTCGCGCA..... 10563
389 rGlyArgMetProPheAlaTrpThrAlaValHisLeuAlaAsnIleVal 405
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10564 .....GCTTACGCCAATCGCT 10581
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGluArg 422
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10582 CCTCGAGACGCGAGCTCCCTTATTCGACCGTCACCGGACACGAGCT 10631
422 gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSer 439
      : |||
10632 CGAC.....GGCTCG 10642
439 eGlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10643 AGCTGACGCGCGCTAGTATCGAACCCTCGGCAACCGCTCG... 10669
456 ValThrAsnPhePheLysGlnGluArgLeuSerAspGluAspLe 472
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10690 .....TTCTGAGCGCGACGAGCGCTCTCGAC...GATGG 10724
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg 487
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10725 GCATCGCTTCTCGTGAAGTCAAGCCCATCCGCTCACGCTCGCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
      |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

10775 TCCGGAGACSTGCGAGGCGTCAACCGCTGATCCCGCTGTCGTCGCC 10824
504 oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10825 ATTCGACGAGAGAGAGGACACCTCGCGCGCTGCTCTCTGCGCGGA 10874
519 LysProTyrrProAspProArgGlyArg..... 527
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
10875 GCTCTCTACCGAGGCTCGCGCTGACTGGAGAGACTTCTTCGCGCCT 10924
528 .....ProThrLysGluIleuGluPheProAla 538
      : |||
10925 ACGTCTCCCGGACAGGCTCTCCCTCCAC.....CTACCCCTTCAGGGA 10968
538 rGluValThrAlaProHisThrSerTyrrArgAsnLeu..... 550
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
10969 GAGCGGTTGCTGCTGAGCTGCCAGAGACAGACGCTTCGACGCTCGCT 11018
551 LeuTyrrValTyrrProHisSerLeuAsnPheSerSerArgGln...GlySer 566
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11019 CCGGAGCGCTGACSTCGCGCGCACCAATCCGCTGTCGCGCGCGCGTGC 11068
567 ValArgAsnLeuAlaValArgValGlnTyrrMetThrGlyLeuAspPro 583
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11069 CTTTCGCGACCGCGGTGGCTTCTCT...TTACAGGCGCGCTCTCCCTC 11115
583 rGln.....AlaLeuPro..... 587
      : |||
11116 GCAGAGACACCGCTGCTGGAAGGCTATCCGCTTCGCGACACCAATCT 11165
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11166 ACCGGGACACCGGCTTCTCGAGCTCGCGCTCGACGTCGCCACCGCGTGC 11215
602 AlaPheThrProValValThrHisAsnLysSerProGluPheTyrrGlu 618
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11216 GCTTCGACACCGTCTC.....GAGA 11234
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11235 GCTACGCTCGAGGCGCTCTCGCTCCCATCGCAGAGACACGCTCTCTC 11283
635 heThrPheTyrrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11284 .....CTCCAG 11289
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg... 666
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11290 ATCTCGTCCGG.....CCGTCGACGACGCGAGACGAGAGGCG 11327
667 .....LeuArgThrGlyProPhe 672
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11328 GCTCTTTCATAGCGGACAGAGAGAGCGGCTTCAGAGTGGCCCTTGA 11377
673 .....CysLeuProVal 676
      : |||
11378 CTCGCCACGCGAGCGCTCTCTCGCGCGGACCCCATCTCTTCGCGC 11427
677 SerValAspGlnProProSerTyrrSerValLeuThrProAspValAl 693
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11428 GATCTCCAGAGGCTCTCTCTCGAGTGCATC.....CCGGTGA 11468
693 aLeuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSer 709
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
11469 CCTCGAAGGCTCTACGCAACCTCGGCAACCTCGGCTTGCCTGCGGCG 11518
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrrLeuAsp 725
      : |||||: ||| ||| ||| ||| ||| ||| ||| ||| |||
11519 CCGAGTTCCAGGCGCTCGCTCTCTACAAAGCGCGCGAGACGCTCTTT 11568
726 LysPhePheThrLeuValHisValLeuGluGlnGlyAlaPheProPhe 742
      : ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
11569 GCCGAGGCAAGCTCCGGAAGCGCGGAGGATCGCGCGGTTTGC 11618

```

742 gReLUys.....AspThrValLeuSerGluGlyAsnValGluG 755
||| |||:
11619 CCTCCACCCCTGGCGTGGCTGCACAGCGCCCTGCATGCACCTGGCTTGAAG 11688
755 InGUleuArgAlaSerLeuAla.....Ala 763
::||| |||:
11669 ACGAGCGAGAGAGGACGGTGCCTGCCTGCCTGAGCGAGCTGC 11718
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisHisValLe 780
|||||: |||:
11719 CTGGCTCCGCTGGGTGCCACCACTTGCCTGCCTGCCTGCAC..... 11760
780 uAspLysLeuValArgLeuValIleArgProProIleIleSerGlyIle 797
||| |||:
11761CGTCCCAAGGGTGAATCCTCCGCT 11785
797 LeValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValValSer 813
:: |||:
11786 CGATCGCTCCTGGCGCGACCGCGAGGTGACCTCTGCTCGGTGCAGAGCG 11835
814 LeuValHisArgSerLeuGluAlaIleArg.....AspAlaArg91 827
|||||: |||:
11836 CTGCGCATGGGAGCGACGTCCGCGCGAGCTCCGACCCCGCGAGCTTC 11885
827 yHisCysProGluLeuAlaAlaIleArgHisTyrAlaPheArgLeuProG 844
||| |||:
11886 CCACCATGATGGCGTC.....TTCGCGGTGCAGCT 11914
844 LyThrGluProSerLeuProAspGlyAlaProProValThrValGlnAla 860
::||| |||:
11915 GAGCGGAGCTCCAAAGCCCTACT..TCACGCGCTGCGCGCCCGAGGCGC 11961
861 AlaThrLeuAlaArgGlySer.....G1 868
:: |||:
11962 GTCCCTCTCGGCACAGCGCGGACGATCTCGGCTGCAGCCCGCTGC 12011
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSerS 884
::||| |||:
12012 CCCTGTACGCGGACCTGCTGCTCCGAAAGCGCCCTGCAGACGAGGCGCTT 12061
884 eArgnProAspLeuAlaValAlaIleProGlySerValAspAspGluValSer 900
|| |||:
12062 CGGCTCCCGGCTGTGTCGCCCC...TTCACTGATCGACCGGCGAGCG 12108
901 ArgIleLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrrPyA 917
::||| |||:
12109 GACCTGTGCCGAGC.....GCCACGAG..... 12132
917 lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrrPheP 934
::||| |||:
12133GCCACGCGCTGCCTGCCTCCCTCTTGA...GCCCTG..... 12168
934 hepHeGlnLeuMetValLysSerMetAlaLeuHisLeuLeuLeuGlyGln 950
|| |||:
12169CTCGCGGACGAA 12180
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
||| |||:
12181 CGCTCGGCTGTGCGGC..... 12198
967 pIleThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH 984
|||||: |||:
12199CTGCTCTCTCAACCGGCGCGCTGCACCGGCGCGCGCGCGCG 12238
984 lAspAspValAlaGluLeuAlaGluHisLeuAsnAlaSerLeuAlaPhePhe 1000
:: |||:
12239 AAGACGACGTCAG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuYa 1017
::||| |||:
12253GACCTCGC 12260

1017 lArgAlaHisTyrLysGluValAlaThrArgLeuGlnSer..SerProAs 1033
::||| |||:
12261 TCACGCCCGGCTGTGGGAGCTCGCGGCTCCGCGCAAGTACACCCAG 12310
1033 nProAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuCys 1050
::||| |||:
12311 ACCTCCGCTCTT.....C 12324
1050 eHisGlnHisTyrValThrLeuAsnLeuProCys..... 1061
::||| |||:
12325 CTGTCGACATCGACCTCAGGAGGCTCCGACAGGCGCTGTAGGCGC 12374
1062CysProLeuSerProPro.....AlaSerProse 1071
||| |||:
12375 GCTGACACAGAGAGACCGAGCTCGGCTCCGCGCAAGGAGAACCTCA 12424
1071 rProSerValSerSerThrThrSerGlnSerSerThrPheSerSerGln 1087
|||||: |||:
12425 TCAGAGGTTGGCGACACACGCTGCAGAGGAGCGCTCATTCGCGCGCA 12474
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyPropheAr 1104
|||||: |||:
12475 GCACCCACGTGGCGCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAla 1121
|||:
12493CATATTCGACCAAGGACCTTCGACGCGCTCGCTCTGCTG 12535
1121 euGluProGluAlaGluGlyAlaPheLeuLeuHisLysAlaIleSer 1137
|||||: |||:
12536 ACGCCCGGAGGCCAG...GCGGCCCTGCACACGCGCAAGTCCGATC 12582
1138 AlaValHisSerLeu..... 1142
||| |||:
12583 GCGGTGCACGCGGAGGCGTCACTCCGCGATGTGTGCACACCTTCG 12632
1143LeuCysGlyHis..... 1146
12633 CATGTATCCGGCGACGCGCGCGCTCGAGAGCGAGAGCGCGGCAATC 12682
1147AspThrAspProArgTyrAlaGluAlaThrValLysAlaArgVal 1161
::||| |||:
12683 TTACTGAGTGTGTCAGGTGTGTCCGATACCGTATGAGGAGCGGATG 12732
1162AlaG1 1163
12733 ATGGGGTCTTCGCGGACCTTGGTCCACGGCCATCGCCGAGCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
::||| |||:
12783 CATGATCTGCCCC.....ATCCCCACGCGCT 12808
1180 lAspPheAlaGluGly..ProGlyGlnArgSerArgLeuAla...SerMe 1195
::||| |||:
12809 GGTCTTCGCCCAAGCGCGAGCGTCCCATCATCTATCATCAGCGCTTAC 12858
1195 tLeuAspSerAspThrGlnGlyGlyGlyAspIleAlaGlyThrIleAsn 1212
::||| |||:
12859 TATGAGCTGTGATCTCGGC.....ATCTGAATACCAATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyGlyProLeuAlaPro 1225
||| |||:
12900 TGTCTCATTCATGGCGCGCGCGGCGGTGCGGAGCGCGCGCGGTTCAGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaG1 1242
::||| |||:
12950 TCGACGCGCACCTCGCGCGCGAGGTCTTGGCACCGCGCATCGAGGAG 12999
1242 yCysAlaLeuSerAlaGlnSerSerArgThrLeuLeuAlaCys..... 1256
::||| |||:
13000 TCGAGGCTCTCGCGCGCTGCGCTTGCAGCATGCGCACCTCGGCTCTC 13049
1257 ..ValLeuTrrValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270

273 GluIleuProIle.....Ph 278
:::|::|
10273 CGCATGCGCCCTGGCGAAGAAAGCGCTCACACCGTCGGCGCAACGG 10322
278 eGlyIleLeuAlaLeuTyrAspValArgIuLysLysIleSerGluA 295
||::|::|
10323 CGGATGGCGCGCTC..... 10338
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
:::|::|::|::|::|::|::|::|::|
10339GACCTGGCGCCCTCCGACTCCAGACTACCTACCTCGCTCC 10377
312 HisGlyThrHisProAlaIleSerThrLeuAla.....ArgSe 324
|||::|::|::|::|::|::|
10378 TGGGGCGACAGG.....CTCTCCAGCGCGCGCTCAACAGCCCCAGGGC 10421
324 rAlaIlePheSerValThrTyrProSerProAspIlePheLeuValIleL 341
:::|::|::|::|::|::|::|
10422 TACCTCTGTATCCGCGAGCCCGCGCTGACAGCGCGCTGCTC..... 10464
341 yslLeuGluLysValLeuGluGlnGlyAspIleSerGluCysGluPro 357
|||::|::|::|::|::|::|
10465GACGTCTCAACCGCCACCAAGGTG 10488
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
:::|::|::|::|::|::|::|
10489 TTCCGCCCGCAAGATCCGCTGACCTCCACTCCGCTCCGCGCAAGAGGA 10538
374 uLysLeuArg.....LeuAlaIleGlnGlnPheCysThrArgLeuGluY 389
:|::|::|::|::|::|::|
10539 CGCCGTCACAGACGCTCCGCCGA..... 10563
389 rGlyArgMetProPheAlaThrThrAlaValHisLeuAlaAsnIleVal 405
|||::|::|::|::|::|::|
10564GCTCTAGCCAACTGCT 10581
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGlnGluArg 422
:::|::|::|::|::|::|::|
10582 CCTCGGAGCTGCGCTCCCTTTATTTCGACCGTCACCGCGCACAGGCT 10631
422 gProAlaThrPheAspArgArgArgGlyProGlnAspArgAlaSerS 439
:::|::|::|::|::|::|::|
10632 CGAC.....GGCTCCG 10642
439 ergIysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
:::|::|::|::|::|::|::|
10643 AGCTGACGCGCGTACTGTATGAAACCTCCGCAACCGCTCTG... 10689
456 ValThrAsnPhePheLysGlnGlnAlaGlnArgLeuSerAspGluAspLe 472
|||::|::|::|::|::|::|
10690TTCTCGAGCGGACGAGCGGCTCTCTCGAC...GATGG 10724
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg. 487
:::|::|::|::|::|::|::|
10725 GCATCGCTTCTCCGTCGAGTCACGCCCATCCGTCGTCAGCTCGGCC 10774
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
|||::|::|::|::|::|::|
10775 TCCGGAGAACCTGAGCGCTCACCGCTCGATCCGTCGTCGTCGCTCC 10824
504 GlnAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
|||::|::|::|::|::|::|
10825 ATTGACGAGAAAGAGCCACCTGCGCCGCTCTCTCTCTGAGCGGA 10874
519 LysProTyrProAspProArgGlyArg..... 527
|||::|::|::|::|::|::|
10875 GCTCTCTACCGAAGCGCTCGGCTCGACTGGAAGACTTCTTGGCGCCT 10924
528ProThrLysGluIleLeuGluPheProAla 538
|||::|::|::|::|::|::|
10925 ACGCTCCCGCAAGAGTCTCTCCCTCCCGAC.....CTACCCCTTCAGGA 10968
538 rgluValLysValAlaProHisThrSerTyrArgAsnLeu..... 550

10969 GAGCGTTCTGGCTCTGAGCTCTCCAGGAGACGCTTCCGAGCTGCCT 11018
551 LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln..GlySer 566
:::|::|::|::|::|::|::|
11019 CGGAGGCTGACTCTGGCCGACCAATCCCGCTGCTGGCGCGCGCGCTCG 11068
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyGluAspProSe 583
|||::|::|::|::|::|::|
11069 CCTTCGCGACGCGGTGGCTTCTCT...TTACAGGGCGGCTCTCTCTCT 11115
583 rGln.....AlaLeuPro..... 587
|||::|::|::|::|::|::|
11116 GCAGAGCACCGCTGGCTCGAAGCCATCCGCTTGGCACACCATCTCT 11165
588ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
|||::|::|::|::|::|::|
11166 ACCGGGACCGGCTTCTGAGCTCGCCCTGAGCTGCCACCGCGCTCG 11215
602 AlaPheThrProValValTyrHisAsnLysSerProGluPheThrGluG 618
|||::|::|::|::|::|::|
11216 GCCTCGACACCGCTC.....GAAGA 11234
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
|||::|::|::|::|::|::|
11235 GCTACGCTCGAGGCGCCCTCTGCTCTCCATCGCAGACACCGCTCTC 11283
635 heThrPheThrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
|||::|::|::|::|::|::|
11284CTCAG 11289
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg..... 666
|||::|::|::|::|::|::|
11290 ANCTCCGCTGG.....CCCGTGAAGAGCGAGGAGGAGG 11327
667LeuArgThrGlyProPhe. 672
11328 GCCTCTTTCATAGCCGACGAGAGAGCGGCTTCAGATGAGCCCTCGA 11377
673CysLeuProVal 676
11378 CTGCGCACGACGCGCTCTCTCTCGCGCGACCCATCTCTCTCGCGC 11427
677 SerValAspGlnProProProSerTyrSerValLeuThrPheAspVal 693
|||::|::|::|::|::|::|
11428 GATCTCACGAGTGCCTCTCCCTGAGTGCATC.....CGGTTGA 11468
693 aleuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709
|||::|::|::|::|::|::|
11469 CTTCGAAGGCTCTACGCAACCTCGCCCAACCTCGGCTTGGCTAGCGCC 11518
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
|||::|::|::|::|::|::|
11519 CCGAGTTCAGGGGCTCTCCCTCTCTACAAAGCGCGCAGCAGACTCTT 11568
726 LysPhePheThrLeuValHisValLeuGluGlnGlyAlaPheProPheAr 742
|||::|::|::|::|::|::|
11569 GCCGAAGCCAGCTCCCGGAGGCGCGGAAAGATGCCGCCGGTTTGC 11618
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGlnG 755
|||::|::|::|::|::|::|
11619 CTTCACACCTGCGCTGCTGACAGACCGCTGATGATGAGCTTGAAG 11668
755 lngLeuAlaArgAlaSerLeuAla.....Ala 763
|||::|::|::|::|::|::|
11669 ACAGAGAGAGAGGAGCGTCTGCTCTGCTCTGCGAGCGAGTCTG 11718
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
|||::|::|::|::|::|::|
11719 CTGGGCTCCGTCGTCGACACCTTGGCGTCCCTCAC..... 11760
780 uAspLysLeuValArgLeuValIleArgProProIleIleSerGlyGlnI 797
|||::|::|::|::|::|::|

11761	CGTCCAAAGGCTGAATCTCTCGTCT	11785
797	LeValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValSer	813	
11786	CGATCTCGTCGGCGGAGCGCGCGAGTGCACCTTGGCTCGGTGAAGG	11835	
814	LeuValHisArgSerLeuGluAlaGln.....	AspAlaArgG1	827
11836	CTGGCCATCGGGAGCGAGCTCCGGCGGAGCTCCGACCCGGCAGCTTC	11885	
827	YHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuProG	844	
11886	CCACCATGATGGCTC.....	TTCCGCGTCGACT	11914
844	LYrhcGluProSerLeuProAspGlyAlaProProValThrValGlnAla	860	
11915	GGAGCGAGCTCCAAACGCCCACT...TCACCGCTGGCGCCCGCAGCGC	11961	
861	AlaThrLeuAlaArgGlySer.....	G1	868
11962	GTCCTTCTGGGACAGGGGGCCAGCATCTCGGCTCGACGCCCGCTCGC	12011	
868	YArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSerS	884	
12012	CCGCTACGGCGAGCTCGCTCGCTCCGAAAGCCCTCGACAGAGGCGCTT	12061	
884	erAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer	900	
12062	CGCGTCCGGCGCTCGTCGTCGCGCCCC...TTATCGATCGACCGCGAGGC	12108	
901	ArgGluLeuAlaSerLysLeuLeuHisGluGluLeuAlaGlnTyrPva	917	
12109	GACCTCGTCCCGACG.....GCCACGAG.....	12132	
917	IValSerSerSerAlaValArgGluAlaIleGluGlnHisAlaTyrPheP	934	
12133GCCACCGGCTCGCAGCTCGCCTCTTGCA...GCTCGG.....	12168	
934	hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGlyGln	950	
12169CTCGCGAGCA	12180	
951	ArgLeuAspThrProArgLysLeuArgPheProGlnArgPheLeuAspAs	967	
12181	CGCCTCGCTCTCGCCG.....	12198	
967	PileThrAlaLeuValGlySerValGlyLeuGluValIleThrArgValH	984	
12199CTCGCTCTGTCACCGCGAGCGCGCTCGCCACCCACAG	12238	
984	IsLysAspValGluLeuAlaGlnIleuAsnAlaSerLeuAlaPhePhe	1000	
12239	AGAGCGAGCTCAG.....	12252	
1001	LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa	1017	
12253GACCTCGC	12260	
1017	ArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSer...SerProAs	1033	
12261	TCACCGCGCGCTCGGGGCTCGCGCGCTCGCGCAAGATGACAGCCAG	12310	
1033	nProAlaIleLeuLeuThrIleuArgMetGluPheThrArgIleLeuCysS	1050	
12311	ACCTCCGCGCTT.....C	12324	
1050	erHisGluHisTyrValThrLeuAsnLeuProCys.....	1061	
12325	CTCGTCGACATCGACCTCAGAGAGCTCCACAGAGCCCTGTAGAGCG	12374	
1062CysProLeuSerProPro.....AlaSerProse	1071	
12375	GCCTGACACAGAGACGCGACGCTCGGCTCGCGAACGGAAACCCCTCA	12424	

1071 p r o s e r v a l s e r s e r t h r s e r l n s e r s e r t h r p h e s e r s e r g l n 1087
12425 T C C G A G G T T G C C A A C C A C A G C G T C A G G A C G C G T C A T C C G C G C A A 12474
1088 A l a p r o a s p p r o l y s v a l t h r s e r m e t p h e g l u l e u s e r c l y p r o p h e a r 1104
12475 G C A C C C C A C G T G C G C C T C 12492
1104 g g l n g l n h i s p h e u n a l a g l y l e u l e u l e t h r g l u l e u a l a l e u a l l 1121
12493 C A T A T T C C G A C C A A A G C A C C T T T G A C G C G C T G C C C T G T G C 12535
1121 e u g l i n p r o l u a l a g l u g l y a l a p h e u l e u h i s t y s l y s a l a l e s e r 1137
12536 A C G C C C C C G A G G C C C A G G C C C C C T C C A C A C G C C C A A G T C C C A T C 12582
1138 A l a v a l h i s s e r l e u 1142
12583 G C C C T G C A C G C G C A G G G C T A C A T T C C G A T G T C T G C A C A C C C T T G G 12632
1143 l e u c y s g l y h i s 1146
12633 C A R T A T C C G G G G G A C C G C G C G C G C T C G G A G C G A A G C G G G C A T C G 12682
1147 A s p t h r a s p r o a r g t y r a l a g l u a l a t h r v a l l y s a l a a r g y a l 1161
12683 T T A C T A G A G T C G G T C C A G A G T G T C C C G A T A C A C C T A G G A C A C G G G T G 12732
1162 A l a g l 1163
12733 A T G G G G T C T T G C G G C A C G C T T T G T C C C A C G G C A T C G C G A C G C G 12782
1163 u l e u t y r l e u p r o l e u s e r t l e a l a a r g a s p t h r l e u p r o a r g l e u h 1180
12783 C A R G A C T G C G C C A T C C C C A C G C C T 12808
1180 l a s p p h e a l a g l u g l y . p r o g l y n a r s e r a r g l e u a l a s e r m e 1195
12809 G G T C T T C G C C A A G C G C C A G C G T C C C A T C A T C T A T C T C A C G C C T A C 12858
1195 t l e u a s p s e r a s p t h r g l u g l y l e u g l y a s p l l e a l a g l y t h r l e a s n p 1212
12859 T A T G G A C T G C T C A T C T C G G C A T C G A A A C C A A T C A C A G 12899
1212 r o s e r v a l a m e t A l a l e a l a g l y l p r o l e u a l a r o 1225
12900 T G T C T C A T C C A T C G C G G C C G C G G C G C T C G G A G G C G C C G T T C A G C 12949
1226 g l y s e r a r g a l a s e r t l e s e r g l n g l y p r o p r o t h r a l a s e r a r g a l a g l 1242
12950 T C C A C A G C C A C C T C G G C G C G A G G T T T G C A C A C G C C A G T C A A G A G A G 12999
1242 y c y s a l a l e u s e r a l a g l u s e r s e r a r g t h r l e u a l a c y s 1256
13000 T G A G G C G T C T C G C G C G T C G G T C G A C A T C G A C A T C G C A C C T G C T C 13049
1257 v a l l e u r v a l l e u s a n t h r g l u p r o a l a l e u c l n 1270
13050 A C G T G A C T G G G C T T C G A C A C A C A C T T C T C G C T C A C A G A T G G C G C G 13099
1271 A r g t y r a l a t h r a s p l e u t h r l e u 1278
13100 G C A T G A T G T C C T C C T C A C T G T C 13123
seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-457-335A-6
seq_documentation_block:
; Sequence 6, Application US/08457335A
; Patient No. 5723759
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Liden, James M.


```

10378 TGGGGCGACAG.....CTCTCCACCGCGCGCTCAACAGCCCGACGGC 10421
324  rAlaIlePheSerValThrTyrProSerProAspIlePheLeuValIleL 341
      ::::: ||| ::::: |||
10422 TACCTCTGTATCCGGCGAGCCCGCCGCTGTGAGCGCGCTGCTC..... 10464
341  yslLeuGluLysValLeuGlnGlnIleLysIleSerGluCysGluPro 357
      ::::: ||| ::::: |||
10465 .....GACGTCCTCACCGCGCCACCAAGCTG 10488
358  TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG1 374
      ::::: ||| ::::: |||
10489 TTCCGCCCGACAGATCCGCGTGTACGCTCCACCTCCGCGCACATGGA 10538
374  ulysLeuArg.....LeuAlaIleGluGlnPheCysThrArgLeuGlyA 389
      ::::: ||| ::::: |||
10539 CGCGGTCCAGAGAGAGCTCCGCCGA..... 10563
389  rGlyTrArgMetProPheAlaTrpThrAlaValHisLeuAlaAsnIleVal 405
      ::::: ||| ::::: |||
10564 .....GCTCAGCCACATCGCT 10581
406  SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGlyLysArgAr 422
      ::::: ||| ::::: |||
10582 CCTCGGACGTGGAGCTCCTTTATTCGACCGTCACCGCGCACAGGCT 10631
422  gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSerS 439
      ::::: ||| ::::: |||
10632 CGAC.....GGCTCCG 10642
439  erGlyAspArgAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
      ::::: ||| ::::: |||
10643 AGCTCGAGCGGCGTACTGTATCGAACCCTCGGCAACCGCTCTG... 10689
456  ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspLeuPle 472
      ::::: ||| ::::: |||
10690 .....TTCTCGAGCGCGAGCGAGCGGCTCTCTGAC...SATGG 10724
472  uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg. 487
      ::::: ||| ::::: |||
10725 GCATCGCTTCTCCGTGAGGTGAGCCGCCATCCGCTGTCACGCGGCC 10774
488  ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaTr 504
      ::::: ||| ::::: |||
10775 TCCGCGAGACTGCGAGCGCTCACCGCTCGATCCGCTGCTGCGGCTCC 10824
504  oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
      ::::: ||| ::::: |||
10825 ATTGAGACGAGAAGAGCCACCTCGCCGCTGCTCTCTGGGCGGGA 10874
519  LysProTyrProAspProArgGlyArg..... 527
      ::::: ||| ::::: |||
10875 GCTCTCTACCCGAGCGCTCGCTCGCTGACGAGGACTTCTTCCGCGCT 10924
528  .....ProThrLysGluIleLeuGlnPheProAlaA 538
      ::::: ||| ::::: |||
10925 AGCGTCCCGGCAAGGTCTCCCTCCAC.....CTACCCCTTCCAGGGA 10968
538  rGluValIlyrAlaProHisThrSerTyrArgAsnLeu..... 550
      ::::: ||| ::::: |||
10969 GAGCGGTCTGCTGCTGCTGCTCCAGGACGAAAGCTTCCGAGCTGGCT 11018
551  LeuTyrValTyrProHisSerLeuAsnPheSerSerArgGln...GlySer 566
      ::::: ||| ::::: |||
11019 CCGCAGCGCTGATCGCGCGACCAATCCGCTCGCGCGCGCGCGCTG 11068
567  ValArgAsnLeuAlaValArgValGlnTyrMetThrGlyLysAspProSe 583
      ::::: ||| ::::: |||
11069 CCTCCCGGACCGCGGCTGCTTCTCT...TTACAGGGGCGGCTCTCCCTC 11115
583  rGln.....AlaLeuPro..... 587
      ::::: ||| ::::: |||
11116 GCAGAGACACCCGTGGCTCGAAGGCCATCGCTCTCGGACACCCATCTCT 11165

```

```

588  .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
      ::::: ||| ::::: |||
11166 ACCGGGACCGCGCTTCTGAGCTCGCCCTGCACGTCCGCCACCGCTCG 11215
602  AlaPheThrProValValTyrHisAsnLysSerProGlnPheTyrGlnG1 618
      ::::: ||| ::::: |||
11216 GCCTCGACACCGCTC.....GAAAGA 11234
618  uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
      ::::: ||| ::::: |||
11235 GCTCACGCTCGAGGCCCTCTGCTCCATCGCAGAGCACCGCTCTC. 11283
635  heThrPheTyrHisValSerCysGlnProAlaArgProGlyThrAlaLeuGlu 651
      ::::: ||| ::::: |||
11284 .....CTCAG 11289
652  ThrProValGlyPheThrTrpIleProLeuGlnHisGlyArg..... 666
      ::::: ||| ::::: |||
11290 ATCTCCGTCGGG.....CCGCTGACGACGACGACGACGAGGC 11327
667  .....LeuArgThrGlyProPhe. 672
      ::::: ||| ::::: |||
11328 GCTCTTTTCCATAGCCGACAAGAGAGCGCTTTCAGATGGCCCTGGA 11377
673  .....CysLeuProVal 676
      ::::: ||| ::::: |||
11378 CTCGCCACCGCAGCGGCTCTCTCGCGCGGACCCCATCCCTCCGCGC 11427
677  SerValAspGlnProProSerTyrSerValLeuThrProAspValAl 693
      ::::: ||| ::::: |||
11428 GATCTCCACAGGTGGCTCCCTCGAGTCCATC.....CGGGTGA 11468
693  aLeuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709
      ::::: ||| ::::: |||
11469 CTTGAAAGCGCTTACGACCAACCTCGCCMACCTCGGCTTCCCTACAGGC 11518
709  aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
      ::::: ||| ::::: |||
11519 CGAGTTTCAGAGGCGTCCGCTCCGCTACAGCCGCGGACGACACTCTT 11568
726  LysPhePheThrLeuValHisValLeuGluGluLysAlaPheProPheAr 742
      ::::: ||| ::::: |||
11569 GCCGAAGCCAACTCCGGAAGCGCGGAAAAGATGACCGCCGCTTTCG 11618
742  gLeuLys.....AspThrValLeuSerGluGlyAsnValGlu 755
      ::::: ||| ::::: |||
11619 CTTCCACCTTCGCTGCTCGACAGCGCCCTGCATGACTGGCTTTGAG 11668
755  IngLuLeuArgAlaSerLeuAla.....Ala 763
      ::::: ||| ::::: |||
11669 ACGAGCAGAGAGGAGAGGTGCTGCTGCTCTCTGAGGAGGAGTCTCG 11718
764  LeuArgLeuAlaSerProGlnProLeuValAlaPheSerHisHisValLe 780
      ::::: ||| ::::: |||
11719 CTGGGCTCTCGGTGGCCACCACTTGGCGGTGCTTCAC..... 11760
780  uAspGlyLeuValArgLeuValIleArgProProIleIleSerGlyGln 797
      ::::: ||| ::::: |||
11761 .....GCTCCCAAGGGTGAATCTCCGCTCT 11785
797  LeValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValSer 813
      ::::: ||| ::::: |||
11786 CGATGCTGCTGGCGGACGCGGAGGTGATACCTCTGTGCTGGGTGCAAGCG 11835
814  LeuValHisArgSerLeuGluAlaIleGln.....AspAlaArgG1 827
      ::::: ||| ::::: |||
11836 CTGCGCATGGGAGAGACTCGCGCGCACGCTCCGACCCCGACACTTC 11885
827  yHisCysProGlnLeuAlaIleTyrValHisTyrAlaPheArgLeuPro 844
      ::::: ||| ::::: |||
11886 CCACCATGATGCGCTC.....TTCCGCGTGAAGCT 11914

```



```
844 1yThrgluProSerLeuProAspGlyAlaProProValThrValGlnAla 860
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
11915 GAGAGGAGTCCAAAGCCCCACT...TCACCGCTGCGCGCCGACGAGC 11961
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
861 AlaThrLeuAlaArgLysSer.....G1 868
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
11962 GTCCTTCGCGCA?AGCGCGCACGATCTCGGCTCGACGCCCGCGCTCGC 12011
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerLysSerLys 884
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12012 CGGCTACGCGCGACCTCGCTCGCTCGCAAGCGCCCTCGACCAAGGCGCTT 12061
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
884 eAsnProAspLeuAlaValAlaProGlySerValAspAspGluValSer 900
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12062 CGCGTCGCGCTCGCTCGCTCGCGCCCC...TTCATGATCGACGCGGACGCC 12108
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
901 ArgLLeuAlaSerLysLeuHisGluGluLeuAlaLeuGlnTrpVal 917
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12109 GACCTGCTCCGAGC.....GCCACGAG..... 12132
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
917 yValSerSerValAlaArgGluAlaLeuGlnHisAlaTrpPhePhe 934
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12133 .....GCCACCGCGCTCGCGACTCGCGCTCTTGCA...GCCG..... 12168
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
934 hePheGlnLeuMetValLysSerMetAlaLeuHisLeuLeuGln 950
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12169 .....CTCGCGCGCGAA 12180
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
951 ArgLeuAspThrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12181 CGCGCTCGCTCGCGC..... 12198
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
967 PheThrAlaLeuValGlySerValGlyLeuGluValLeuThrArgValH 984
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12199 .....CTCGCTCGCTCGCGCGCGCGCGCTCGCGCACACACACCG 12238
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
984 isLysAspArgLLeuLeuAlaGlnHisLeuAsnAlaSerLeuAlaPhePhe 1000
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12239 AAGACGACGCTCAAG..... 12252
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12253 .....GACCTCGC 12260
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1017 yArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSer...SerProAs 1033
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12261 TCACGCGCGCTCGGCGGCTCGCGCTCGCGCAAAAGTGAAGCAACCGAG 12310
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1033 nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgLLeuGln 1050
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12311 ACCTCCGCTCT.....C 12324
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1050 eThHisGlnHisTyrValThrLeuAsnLeuProGln..... 1061
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12325 CTGCTGACATGACCTGACGCGGCTCGCGACGAGGCTCGTAGAGCGC 12374
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12375 GCTCGACACAGAGAAAGCGGACCTCGCGCTCGCGAAAGGAAACCGCTCA 12424
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1071 rProSerValSerSerThrThr.SerGlnSerSerThrPheSerSerGln 1087
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12425 TCSCGAGGTGGCGGACACGACGCTCGAGCGAGCGCTCATCGCGCGCA 12474
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1088 AlaProAspProLysValThrSerMetPheGlnLeuSerGlyProPheArg 1104
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12475 GACGCCACGCTGGCGCTC..... 12492
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAla 1121
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12493 .....CATATTCGACCAAGGACCTTCGACGCGCTCGCGCTCGCTG 12535
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1121 eugLurProGluAlaGlnGluAlaPheLeuLeuHisLysValAlaLeuSer 1137
```

```
12536 ACCGCCCGGAGGCCAG...GGCGCCCTCGCACACGCGCAAGTCCGATC 12582
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1138 AlaValHisSerLeu..... 1142
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12583 GCCGTGACAGCGGCGAGGGCTCAACTTCGCGATGTGTCGACACCTTGG 12632
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1143 .....LeuGlyHis..... 1146
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12633 CATGATCCGGGCGAGCGCGCGCTCGAGAGCGAGAGCGCGGCGATCG 12682
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1147 .....AspThrAspProArgTyrAlaGlnAlaThrValLysAlaArgVal 1161
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12683 TTAGTAAGTGTGTCAGAGTGTCTCCGATACCGCTAGGCGAGCGGGTG 12732
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1162 .....AlaG1 1163
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12733 ATGGGGGTCTTGCGCGACGCTTGCTGTCACGGCATCGCGGACGCGCG 12782
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1163 uLeuTyrLeuProLeuLeuSerLLeuAlaArgAspThrLeuProArgLeuH 1180
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12783 CATGATCTGCGCC.....ATCCCGACGCGCT 12808
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1180 isAspPheAlaGluGly.ProGlyGlnArgSerArgLeuAla...SerMe 1195
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12809 GGTCTTCGCGCGACGCGCGACGCTCCCATCATCATCTACATCGCGCTAC 12858
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1195 tLeuAspSerAspThrGlnGlyGluGlyAspLLeuAlaGlnP 1212
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12859 TATGAGCTGTGATCTCGGCG.....ATCGAAGACCATTCACAG 12899
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1212 roSerValAlaMet.....AlaLLeuAlaGlyGlyProLeuAlaPro 1225
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12900 TGGGCTCATCATAGCGCGCGCGCGCGCTCGAGCGCGCGCGCTCGAGC 12949
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1226 GlySerArgAlaSerLLeuSerGlnGlyProProThrAlaSerArgAlaG1 1242
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
12950 TCGACGCGCACCTCGGCGCGCGAGGTCTTGGCACCGCGAGTCCAGGAGAG 12999
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1242 yCysAlaLeuSerAlaGlnLeuSerArgTThrLeuLeuAlaCys..... 1256
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13000 TGGAGCGCTCTCGCGCGCGCTCGGCTTCGACGATCGCGACCTCGGCTCTC 13049
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1257 ..ValLeuThrValLeuLysAsnThrGln.....ProAlaLeuLeuGln 1270
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13050 ACCTGACCTGGGCTTCGAGACGACCTTCCTCGCGCTCGACGATGGCGCG 13099
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1271 ArgTyrAlaThrAspLeuThrLeu 1278
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
13100 GCATGATGTGCTCGCTCGACTGTG 13123
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-729-214-6
seq_documentation_block:
; Sequence 6, Application US/08729214
; Patent No. 5817502
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight Steven
; APPLICANT: Ryals, John Andrew
; APPLICANT: Hammer, Phillip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kliner, Sabine
; TITLE OF INVENTION: Genes for the synthesis of
; TITLE OF INVENTION: antipathogenic substances
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESS: Ciba-Geigy Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
```



```

406 SerSerAlaGlyLeuAspArgSerSerGluGluArgAr 422
    : : : : : : : : : : : : : : : : : :
10582 CCGGAGCGTCTCGTCTTTATTCAGCCGACGGCGCT 10631
    : : : : : : : : : : : : : : : : : :
422 gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSerS 439
    : : : : : : : : : : : : : : : : : :
10632 CGAC.....GGCTCCG 10642
    : : : : : : : : : : : : : : : : : :
439 erGlyAspArgAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
    : : : : : : : : : : : : : : : : : :
10643 AGCTCGACGGCGCTACTGCTATCGAAACCTCGGCAACCGCTCTG... 10689
    : : : : : : : : : : : : : : : : : :
456 ValThrAsnPhenPheGlyGlnGluAlaArgLeuSerAspGluSple 472
    : : : : : : : : : : : : : : : : : :
10690 .....TTCTCGAGCGCGACCGCGGCTCTCTCGAC...GATCG 10724
    : : : : : : : : : : : : : : : : : :
472 uphLeuPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg. 487
    : : : : : : : : : : : : : : : : : :
10725 GCATCGGCTTCGTCGAGGTGAGCCGCCATCCCGTCTCAGCTCGCC 10774
    : : : : : : : : : : : : : : : : : :
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaPr 504
    : : : : : : : : : : : : : : : : : :
10775 TCCGCGAGACCTGCGAGCGCTCACCGCTGATCCGCTCGTCGCGCTCC 10824
    : : : : : : : : : : : : : : : : : :
504 ogLuanProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
    : : : : : : : : : : : : : : : : : :
10825 ATTCGACGAGAAAGAGCCACCTCGCCCGCTGCTCTCTCTCGGCGGA 10874
    : : : : : : : : : : : : : : : : : :
519 LysProLysProAspProArgGlyArg..... 527
    : : : : : : : : : : : : : : : : : :
10875 GCTCTGTACCCGAGCGCTCGCGCTGACGTGAAGACCTCTTCGCGCCCT 10924
    : : : : : : : : : : : : : : : : : :
528 .....ProThrLysGluIleLeuGlnPheProAla 538
    : : : : : : : : : : : : : : : : : :
10925 ACCGCTCCCGCAAGCTCTCCCTCCAC.....CTACCCCTTCACGGA 10968
    : : : : : : : : : : : : : : : : : :
538 rgGluValTyrAlaProHisThrSerTyrArgAsnLeu..... 550
    : : : : : : : : : : : : : : : : : :
10969 GACGCGTCTGCTGCGACGTCTCCACGACGACGACGCTTCGACGTCGCT 11018
    : : : : : : : : : : : : : : : : : :
551 LeuTyrValTyrProHisSerLeuAsnPheserSerArgGln..GlySer 566
    : : : : : : : : : : : : : : : : : :
11019 CCGCAGGCTGACCTCGCGACGACCACTCCGCTCGCGCGCGCGCTCG 11068
    : : : : : : : : : : : : : : : : : :
567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGluGluAspProSe 583
    : : : : : : : : : : : : : : : : : :
11069 CCGTCGCCGACCGCGCTGCTTCTCT...TTACAGGCGGCTCTCTCC 11115
    : : : : : : : : : : : : : : : : : :
583 rGln.....AlaLeuPro..... 587
    : : : : : : : : : : : : : : : : : :
11116 GCAGAGCACCCGTGCTCGAAGGCATGCGTCTCGGCACACCATCT 11165
    : : : : : : : : : : : : : : : : : :
588 .....ValIlePheGlyLysSerSerCysSerGluPheThrArgGlu 601
    : : : : : : : : : : : : : : : : : :
11166 ACCGGGACCGGCTTCTCTGACCTCGCCCTCGACGTCGCCACCGCGTGC 11215
    : : : : : : : : : : : : : : : : : :
602 AlaPheThrProValValTyrHisAsnLysSerProGluPheTyrGluG 618
    : : : : : : : : : : : : : : : : : :
11216 GGCTCGACACCGTC.....GAGA 11234
    : : : : : : : : : : : : : : : : : :
618 uphLeuLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
    : : : : : : : : : : : : : : : : : :
11235 GCTCAGGCTCGAGGCCCTCTGCTCCATCGCAGACGACCGTCTCTC 11283
    : : : : : : : : : : : : : : : : : :
635 heThrPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
    : : : : : : : : : : : : : : : : : :
11284 .....CTCCAG 11289
    : : : : : : : : : : : : : : : : : :
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg..... 666
    : : : : : : : : : : : : : : : : : :
11290 AATCGCGTGGG.....CCGCTGAGCAGCAGACGACGAGAGGCG 11327
    : : : : : : : : : : : : : : : : : :
667 .....LeuArgThrGlyProPhe. 672
    : : : : : : : : : : : : : : : : : :

```

```

11328 GCTCTTTTCATACCGCAAGAGACGCGCTTCAGATGACCGCCCTGGA 11377
    : : : : : : : : : : : : : : : : : :
673 .....CysLeuProVal 676
    : : : : : : : : : : : : : : : : : :
11378 CTCGCCAGCGCAGCGGCTCTCTCGCGCGACCCCAATCCCTCTCCGCC 11427
    : : : : : : : : : : : : : : : : : :
677 SerValAspGlnProProProSerTyrSerValLeuThrProAspValAl 693
    : : : : : : : : : : : : : : : : : :
11428 GATCTCCAGATGGCTCCCTCGAGTGCATC.....CCGCTGGA 11468
    : : : : : : : : : : : : : : : : : :
693 aLeuProGlyMetLagTyrValAspGlyHisLysGlyVal...PheSerV 709
    : : : : : : : : : : : : : : : : : :
11469 CCTCGAAGGCTCTACGACACCTCGCACCTCGGCTGCTCTACGCGCC 11518
    : : : : : : : : : : : : : : : : : :
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeuAsp 725
    : : : : : : : : : : : : : : : : : :
11519 CCGAGTTCCAGGCGCTCGCTCGCTCTACAAAGCGGCGACGACGCTCTT 11568
    : : : : : : : : : : : : : : : : : :
726 LysPhePheThrLeuValHisValLeuGluGluAlaPheProPheAr 742
    : : : : : : : : : : : : : : : : : :
11569 GCCGAAGCCAGCTCCGGAAGCGCGCAAAAGATGCCGCGGTTTGC 11618
    : : : : : : : : : : : : : : : : : :
742 gLeuLys.....AspThrValLeuSerGluGluAsnValGluG 755
    : : : : : : : : : : : : : : : : : :
11619 CCTCCACCCCTCGCTGCTGACAGCGCTCGACATGAGCTTTGAGG 11668
    : : : : : : : : : : : : : : : : : :
755 IngLueuArgAlaSerLeuAla.....Ala 763
    : : : : : : : : : : : : : : : : : :
11669 ACGACAGAGAGGAGCGCTGCTGCGCTCTGCTGAGAGCGAGTCTCG 11718
    : : : : : : : : : : : : : : : : : :
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
    : : : : : : : : : : : : : : : : : :
11719 CTGCGCTCGCTGCGGACCACTTGGCGCTGCTCCAC..... 11760
    : : : : : : : : : : : : : : : : : :
780 waSprLysLeuValArgLeuValIleArgProIleIleSerGlyGlnI 797
    : : : : : : : : : : : : : : : : : :
11761 .....GCTCCAGGCTGAACTCTCCCTCT 11785
    : : : : : : : : : : : : : : : : : :
797 leValAsnLeuGlyArgGlyAlaPheGluAlaMetAlaHisValValSer 813
    : : : : : : : : : : : : : : : : : :
11786 CGATGCTGCTGCGCGAGCGCGAGTGACCTCTGCTGCTGCTCAAGCG 11835
    : : : : : : : : : : : : : : : : : :
814 LeuValHisArgSerLeuGluAlaAlaGln.....AspAlaArgG 827
    : : : : : : : : : : : : : : : : : :
11836 CTGCGCATGCGAGCAGCTCGCGCGACGCTCCGACCCCGGACCTCT 11885
    : : : : : : : : : : : : : : : : : :
827 yHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPheArgLeuProG 844
    : : : : : : : : : : : : : : : : : :
11886 CCACCATGATCGCTC.....TTCCGCGTCACT 11914
    : : : : : : : : : : : : : : : : : :
844 LyrThrGluProSerLeuProAspGlyAlaProProAlaThrValAlaAla 860
    : : : : : : : : : : : : : : : : : :
11915 GAGACGAGCTCAAGACCCCACT...TCACCGCTGCGCGCGGAGGCG 11961
    : : : : : : : : : : : : : : : : : :
861 AlaThrLeuAlaArgGlySer.....G 868
    : : : : : : : : : : : : : : : : : :
11962 GTCCCTTCGCGACAGCGCGACGATCTCGGCTCGACCCCGCTCGC 12011
    : : : : : : : : : : : : : : : : : :
868 yArgProAlaSerLeuTyrLeuAlaArgSerLys...SerIleSerSers 884
    : : : : : : : : : : : : : : : : : :
12012 CCGGTACGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12061
    : : : : : : : : : : : : : : : : : :
884 erAsnProAspLeuAlaValAlaProGlySerValAspArgGluValSer 900
    : : : : : : : : : : : : : : : : : :
12062 CCGCTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 12108
    : : : : : : : : : : : : : : : : : :
901 ArgThrLeuAlaSerLysLeuLeuHisGluGluLeuAlaLeuGlnTrpVa 917
    : : : : : : : : : : : : : : : : : :
12109 GACTGTGCGCGAGC.....GCCACGAG..... 12132
    : : : : : : : : : : : : : : : : : :
917 lValSerSerSerAlaValArgGluAlaIleLeuGlnHisAlaTrpPhe 934
    : : : : : : : : : : : : : : : : : :

```

```

12133 .....GCCACCGCGCTCGACCTGCGCCCTCTTGCAA...GCCTGG.... 12168
934 hepheglnleumetvalylssermetalaaleuhsleuileuglyln 950
12169 .....CTCGCCGACGAA 12180
951 ArgLeuAspThrProArgTylsLeuArgPheProGlyArgPheLeuAspAs 967
12181 CGCCTCGCCTCGCGCC..... 12198
967 pilerhalaaleuvalglyservalglyleuGluValIlethrArgValH 984
12199 .....CTGCTCTCTGTCACCCGACGCGCTCGCCACCCACCCACCG 12238
984 lslYsAspValGluLeuAlaGluHlsLeuAsnAlaSerLeuAlaPhehe 1000
12239 AAGACACGCTCAAG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVa 1017
12253 .....GACCTCGC 12260
1017 lArgAlaHlsTyrLysGlnValAlaIleThrArgLeuGlnSer..SerProAs 1033
12261 TCACGCGCGCGCTCTGGGGGCTCGCGCGCTCCGCAAGTGACGACCCAG 12310
1033 nProAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuGlyss 1050
12311 ACCTCCGCTCTT.....C 12324
1050 exHlsGlnHlsTyrValThrLeuAsnLeuProGys..... 1061
12325 CTGCTGACATGACACTCGACGCGCTCCGACAGCCCTGTAAGCGC 12374
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
12375 GCTCGACACAGAGAAAGCCAGCTCGCTCCGCAAGCGGAAACCTCTA 12424
1071 rProSerValSerSerThrThr..SerGlnSerSerThrPheSerGln 1087
12425 TCCCGAGGTTGGCGCAACCAAGCTCGACGAGCGCTCATCCCGCGCAA 12474
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyProPheAr 1104
12475 GCACCCACGAGGCGCTC..... 12492
1104 gGlnGlnHlsPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAla 1121
12493 .....CATATTCCGACCAAGGACCTTCGACGCGCTCGCTGTCG 12535
1121 euGluProGluAlaGluGlyAlaPheLeuLeuHlsLysAlaIleSer 1137
12536 ACCGCCCCGAGGCCAG...GGCCCCCTGCGACACGCGCAAGTCGCGC 12582
1138 AlaValHlsSerLeu..... 1142
12583 GCGCTGACGCGGACGAGGCTCAACTCCGCGATGTCGTGACACCTTG 12632
1143 .....LeuGysGlyHls..... 1146
12633 CATGTATCCGGGCGACGCGCGCGCTCGAGGCGGAAAGCGCGGCGCATG 12682
1147 .....AspThrAspProArgTyrAlaGluAlaThrValAlaIleArgVal 1161
12683 TTTACTGAAGTCGTCGAGGTCTCCCGATACACCTAGCGACGCGGTG 12732
1162 .....AlaGln 1163
12733 ATGGGGGTTTCGGCGACGCTTGGTCCACGCGCATCGCGACGCGCGCG 12782
1163 ulLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTCGCC.....ATCCGCCACGCGCT 12808

```

```

1180 lAspPheAlaGluGly..ProGlyGlnArgSerArgLeuAla...SerMe 1195
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12809 GGTCTTCGCCCAAGCGCCAGCGGTCCCATCATCATCATGACCGGCTAG 12858
1195 lLeuAspSerAspThrGluGlyGlnGlyAspIleAlaGlyThrIleAsp 1212
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12859 TATGACTCTCGATCTCGGGC.....ATCTGAACCAATCATCAG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyGlyProLeuAlaPro 1225
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12900 TGTCCTCATCGATCGGCGCGCGCGCGCTGGGACGCGCCGCTTCAGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAla 1242
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
12950 TCGCAGCCCACTCGGCGCGAGGTCTTGCCACCGCACTCCAGGAGAG 12999
1242 yCysAlaLeuSerAlaGluSerSerArgThrLeuLeuAlaCys..... 1256
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
13000 TGGAGCGCTCTCCGCGCTCGGCTTCGACGATGCGCACCTCGCTCTC 13049
1257 ..ValLeuTrpValLeuLysAsnThrGlu.....ProAlaLeuGln 1270
13050 ACGTGACCTGGGCTTCGAGCAGCACTTCTGCGCTCCACCATGGCGCG 13099
1271 ArgTrpAlaThrAspLeuThrLeu 1278
13100 GCATGAGATGCTCTCGACTGTC 13123

seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:us-09-028-934-6

seq_documentation_block:
; Sequence 6, Application us/09028934
; Patent No. 6117670
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Lam, Steven T.
; APPLICANT: Hammer, Philip E.
; APPLICANT: van Pee, Karl-Heinz
; APPLICANT: Kirner, Sabine
; APPLICANT: Young, Thomas R.
; TITLE OF INVENTION: Pyroclitric Biosynthesis Genes and Uses
; TITLE OF INVENTION: Theoret
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 6117670artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: NC
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/028,934
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC1506/CIP7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587

```

```
TELEFAX: 919-541-8689
;
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28958 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; US-09-028-934-6

alignment_scores:
    Quality: 221.00      Length: 1480
    Ratio: 0.350         Gaps: 78
    Percent Similarity: 42.703   Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-09-028-934-6 ..

Align seg 1/1 to: US-09-028-934-6 from: 1 to: 28958

16 ThrValAlaAlaGluValAlaArgLysGluValSerArgGluArgSerGlyse 32
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
9487 ACACCTGCGCATCGACGCTC.....GTTGATACCCCGAGGCTCTTGCGGC 9527
   32 rProHis.....SerSerArgArgCysSer...SerSerLeuGlyV 45
   : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9528 CCGCCACGAGATATACATCCCGCCGCGCGCGCTCTCCGCTTGCGAGC 9577
   45 aLProLeuThrGluValValGluProLeuAspPheGluAspValLeuLeu 61
   :|: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9578 TCtCCGCGCACCAACGCC.....CAGCTCATCTTC 9606
   62 SerArgProProAspAlaGluProGlyProLeuArgAspLeuValGluPh 78
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9607 GAGGAGGCTCCCGCGCGCTGTCGCGGC..... 9633
   78 eProAlaAspAspLeuGluLeuLeuLeuGluProArgGluCysArgThr 95
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9634 .....GAGCCCGCCACCTCACAAGACGG 9655
   95 hrGluProGlyLeuProLys.....AspGluLysLeu 105
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9656 CGTGCGGACCGCTCCCGCGCGCTGCCGTCTCTGTGCGGCAAGAC 9705
   106 AspAlaGluValArgAlaAlaValGluMetTyrTleGluAspTyrValTl 122
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9706 GAGGCGCGCTCGCGCGCGGAGGAGCG...CTCCGCGACCACTCTCT 9752
   122 eValHisArgArgTyrGlnTyrLeuSerAlaAlaTyrSerProValThrT 139
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9753 CGCCCAAGACGACCTCGCTTATCGATGCGTATTCAGAGCCACCA 9802
   139 hrAspThrGln.....ArgGluArgGln 146
   ||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9803 CCGCGCCGCCACTTCGAGCACCGCGCGCTCTCTGTGCGCGGACCGCAC 9852
   147 LysGlyLeu.ProArgGluVal..... 153
   :||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9853 GAGCTCTCTCGCGCTCGCTGCGCCAGAGCAAGCCGCCCGCCGAG 9902
   154 .....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9903 CACGGTTCGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 9952
   169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
   :||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
9953 CTGGGCAAGGCTCGAGTGGGAAGGATGCGCCCTCTCTGCTGCACTCC 10002
   183 oArgSerSerGlyAlaSerSerTlePheAspLeuArgAsnLeuAlaAlaA 200
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
```

```
10003 TCGCCGGCTCTTCGCGCTCAGC.....TCGAGCATCCGAGCGCGC 10043
200 spSerLeuLeu.....ProSerLeuGluValArg 209
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10044 GCTGCTCTCTACGTCGAGTGAGGCTCTGCGCTGCGCTGCGCGGAGCG 10093
210 AlaAlaProGluAspValAlaSpArgArgAsnGluThrLeuArgGlnHl 226
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10094 AAGGCGCCCTCC.....TCGACCGCTGAGCTGCTA 10128
226 sArgProProAlaLeuThrLeuThrProAlaProAspGluAspGlu. 242
   : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10129 CAGCCGCTCTTGTGCGCTAGCTCTGCGCGCTGCGCTGCGCTGCGCTC 10178
243 Ala.....ValGluArgCysSerArgProGluProProArgGluHl 256
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10179 GCTGCGCTGAGCGCGCGCGCTGCTGCGCGCACAGCGGCGAGATCG 10228
256 sPheGlyGlnArgTle.LeuValLysCysLeuSerLeuLysPheGluTle 272
   : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10229 C.....CGCGCGCTTCGTCGAGCGCTCTCTCTGAGAGCGCGCGC 10272
273 GluTleGluProIle.....Ph 278
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10273 CGCATCGCGCGCTCGCAGAGAGCTGACACCGTCCGCGGCAAGCG 10322
278 eGlyTleLeuAlaLeuTyrAspValArgLysLysLysLysLysLysLys 295
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10323 CGCGATGCGCGCGCTC..... 10338
295 snPheTyrPheAspLeuAsnSerAspSerMetLysGlyLeuLeuArgAla 311
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10339 .....GAGCTCGCGCTCCGACCTCAGACCTACCTCTGCTCC 10377
312 HisGlyThrHisProAlaIleSerThrLeuAla.....Argse 324
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10378 TGGGCGCAGAG.....CTCTCCACGCGCGCGTCAACAGCCCAAGGCG 10421
324 rAlaIlePheSerValThrTyrProSerProAspIlePheLeuValIleL 341
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10422 TACCTCTGATTCGCGAGCGCGCGCTGCGAGCGCTGCTC..... 10464
341 yLLeuGluLysValLeuGlnGlnGlyAspIleSerGluCysGluPro 357
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10465 .....GAGCTCTCACCAGCCACCAAGGTG 10488
358 TyrMetValLeuLysGluValAspThrAlaLysAsnLysGluLysLeuG 374
   :||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10489 TTCGCGCCGAGATCCGCGCTGAGCTGCTCCACCTCCGCGGCAAGTGA 10538
374 uLysLeuArg.....LeuAlaAlaGluGlnPheCysThrArgLeuGlyA 389
   : |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10539 CGCGCTCCAGACGAGCTGCGCGCA..... 10563
389 rGlyArgMetProPheAlaThrPheAlaValHisLeuAlaAsnIleVal 405
   |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10564 .....GCTTACGACATCGCT 10581
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGluArgAr 422
   :||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10582 CCGGAGAGCTGAGAGCTCCCTTATTGACAGCTCACCGGACCAAGGCT 10631
422 gProAlaThrPheAspArgArgArgGlyProGlnAspArgAlaSerS 439
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10632 CGAC.....GGCTCCG 10642
439 eArgLysAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
   :||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10643 AGCTTCGAGCGCGCTACTGATGAAACCTCCGGAACAGCTCTC... 10689
456 ValThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLe 472
   :|||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
10690 .....TTCGAGGCGGACCGAGGCGCTCTCGAC...GATGG 10724
```



```

12239 AAGACGAGCTCAAG..... 12252
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017
12253 .....GACCTCGC 12260
1017 ArgAlaHisTyrLysGlnValAlaThrArgLeuGlnSer..SerProAs 1033
12261 TCACGCGCCGCTCTGGGGCTCGCGCGCTCCGCGCAAGTGCACACCCAG 12310
1033 nPrAlaAlaLeuLeuThrLeuArgMetGluPheThrArgIleLeuGlyss 1050
12311 ACCTCCGCGCTCT.....C 12324
1050 eRhISGlnHisTyrValThrLeuAsnLeuProCys..... 1061
12325 CTGCTGCACATCGACCTCAGAGAGCGCTCCACAGAGCCCTGCTAGGCGC 12374
1062 .....CysProLeuSerProPro.....AlaSerProse 1071
12375 GCTCGACACAGAGACAGCCAGCTCTCCGACAGGGAGAACCCCTCA 12424
1071 rProSerValSerSerThrThr..SerGlnSerSerThrPheSerSerGln 1087
12425 TCCTCGAGGTGGCCCAACACAGCTCGACGAGCGCTCATCCGCCGCA 12474
1088 AlaProAspProLysValThrSerMetPheGluLeuSerGlyPropheAr 1104
12475 GCACCCACAGTGGCGCTC..... 12492
1104 gGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAla 1121
12493 .....CATATTCGACCAAGGACCTTCGACGCGCTCGCTCGCTC 12535
1121 euGluProGlnAlaGlnGlyAlaPheLeuLeuHisLysLysAlaIleSer 1137
12536 ACGCCCCCGAGGCCAG...GCCGCCCTCGACACGCGCAAGTCCGCATC 12582
1138 AlaValHisSerLeu..... 1142
12583 GCCGTGCACGCGGAGGGCTCACTTCGCGATGCTGCACACCTTGG 12632
1143 .....LeuCysGlyHis..... 1146
12633 CATGATTCGGGGCAGCGCGCGCTCGAGGGAGCGCGGCATCG 12682
1147 .....AspThrAspProArgTyrAlaGlnAlaThrValLysAlaArgVal 1161
12683 TTACTGAAGTGGTCCAGGTGTCTCCGATACACGTAAGCGGACCGGTG 12732
1162 .....AlaGln 1163
12733 ATGGGGGTCTTCGGCGCAGCTTTGGTCCACGCGCATCGCGGAGCGCG 12782
1163 uLeuTyrLeuProLeuLeuSerIleAlaArgAspThrLeuProArgLeuH 1180
12783 CATGATCTGCCCC.....ATCCCAACGCGCT 12808
1180 IsAspPheAlaGlnGly..ProGlyGlnArgSerArgLeuAla...SerMe 1195
12809 GGTCTTCGCCCAAGCGCGCGCTCCCATCTCATCTACCGCGCTAC 12858
1195 tLeuAspSerAspThrGlnGlyGlyAspIleAlaGlyThrIleAsn 1212
12859 TATGACTCTGCTCATCTCGGC.....ATCTGAACCAACCAATCAACG 12899
1212 roSerValAlaMet.....AlaIleAlaGlyGlyProLeuAlaPro 1225
12900 TGTCTCTACATCGATGCGCGCGCGGTGCGGCGCGCGCGCTTACGC 12949
1226 GlySerArgAlaSerIleSerGlnGlyProProThrAlaSerArgAlaGln 1242

```

```

12950 TCGCAGCCACTCGGCGCGCGAGTCTTTGCCACCGCCAGTCCAGGAG 12999
1242 yCysAlaLeuSerAlaGlnSerSerArgThrLeuLeuAlaCys..... 1256
13000 TGGAGCGCTCTCCGCGCGCTCGCTTCGACAGTGCAGCATCGCGCTCTC 13049
1257 ..ValLeuTyrValLeuLysAsnThrGlu.....ProAlaLeuLeuGln 1270
13050 AGGTGACCTGGGCTTCGACACAGCACTTCCTGCGCTCCAGCATGGGCGCG 13099
1271 ArgTyrAlaThrAspLeuThrLeu 1278
13100 GCATGATGTCTGCTTCGACTGTC 13123

seq_name: /cgn2_6/plodata/2/ina/5A_COMB.seq:US-08-764-233A-1
seq_documentation_block:
: Sequence 1, Application US/08764233A
: Patent No. 5716849
: GENERAL INFORMATION:
: APPLICANT: Ligon, James M.
: APPLICANT: Schupp, Thomas
: APPLICANT: Beck, James J.
: APPLICANT: Hill, Dwight S.
: APPLICANT: Neff, Snezana
: APPLICANT: Ryals, John A.
: TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Ciba-Geigy Corporation
: STREET: 520 White Plains Road, P.O. Box 2005
: CITY: Tarrytown
: STATE: NY
: COUNTRY: USA
: ZIP: 10591
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/764,233A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/729,214
: FILING DATE: 09-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/258,261
: FILING DATE: 08-JUN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: 1506/CIP6
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 49377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Sorangium cellulosum
: IMMEDIATE SOURCE:
: CLONE: p98/1, p1J3, and pVKM15
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 383..760
: OTHER INFORMATION: /product="SorR"
: OTHER INFORMATION: /note="This gene encodes a protein that is highly homolog

```



```
OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
OTHER INFORMATION: Saccharopolyspora erythraea."
FEATURE:
NAME/KEY: misc_feature
LOCATION: 927..19874
OTHER INFORMATION: /product= "Sora"
OTHER INFORMATION: /note= "gene product is highly homologous to type I PKs that
OTHER INFORMATION: are known to be involved in the synthesis of polyketide
FEATURE:
NAME/KEY: misc_feature
LOCATION: 942..7115
OTHER INFORMATION: /product= "Module 1 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7203..12884
OTHER INFORMATION: /product= "Module 2 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 13455..19616
OTHER INFORMATION: /product= "Module 3 of Sora"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19871..46318
OTHER INFORMATION: /product= "Sorb"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19870..24556
OTHER INFORMATION: /product= "Module 1 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 24638..30820
OTHER INFORMATION: /product= "Module 2 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 30881..35446
OTHER INFORMATION: /product= "Module 3 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 35528..40114
OTHER INFORMATION: /product= "Module 4 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 40190..46318
OTHER INFORMATION: /product= "Module 5 of SorB"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 46851..47891
OTHER INFORMATION: /product= "Sorm"
OTHER INFORMATION: /note= "the protein encoded by the sorm gene is highly
OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
OTHER INFORMATION: hygroscopicus that is involved in the synthesis of the
OTHER INFORMATION: polyketide rapamycin."
US-08-764-233a-1

alignment_scores:
Quality: 221.00      Length: 1480
Ratio: 0.350         Gaps: 78
Percent Similarity: 42.703 Percent Identity: 22.500

alignment_block:
US-09-736-968a-2 x US-08-764-233a-1 ..
Align seg 1/1 to: US-08-764-233a-1 from: 1 to: 49377

16 ThValAlaAlaGluValArgGlnValSerArgGluArgSerGlyse 32
||||:|||||:|||||:|||||:|||||:
25790 ACACGTGCGCCATCGACGTC.....GTTGATACCCGACGAGCTTGCGC 25830
32 rProHis.....SerSerArgArgCysSer...SerSerLeuGlyV 45
|||:|||||:|||||:|||||:|||||:
```

```
25831 CCGCCAGCAAGATAGCATGCCCGCGCGCGCGGCTCTCCGCTTCGGAC 25880
45 aLProLeuThrGluValAlaGluProLeuAspPheGluAspValLeuLeu 61
::|:|||||:
25881 TCTCCGCGCACCAAGCC.....CACGTCACTCTC 25909
62 SerArgProProAlaGluProGlyProLeuArgAspLeuValGluPh 78
::|:|||||:
25910 GAGAGAGCTCCCGCCGCTCTGCGGC..... 25936
78 eProAlaAspAspLeuGluLeuLeuGluProArgGluCysArgThrT 95
::|:|||||:
25937 .....GAGCCGCCACCTCATCAGACGG 25958
95 hrGluProGlyLeuProlys.....AspGluLeu 105
::|:|||||:
25959 CGTCGCGACCGCTCCCGCGCGGTGCTGCCGTCTCTGCGCGGAGC 26008
106 AspAlaGluValArgAlaValAlaGluMetTyrIleGluAspTrpValI 122
::|:|||||:
26009 GAGCGCGCGCTCGCGCGCGCGGAGGAGG...CTCCGAGACACTCTCT 26055
122 eValHisArgArgTyrGlnTyrLeuSerAlaAlaTyrSerProValThr 139
::|:|||||:
26056 CCGCCAGCAGCACTCGCCCTTATGATGTGGCTTATTCGAGGCGCACCA 26105
139 hrAspThrGln.....ArgGluArgGln 146
||:|||||:
26106 CCGCGCGCCACTTCAGACCGCGCGCTCTCTCTGCGCGGACCGCGAC 26155
147 LysGlyLeu.ProArgGlnVal..... 153
::|:|||||:
26156 GAGCTCTCTCTCCGCTCGACTCGCTCGCCAGACAGACCGCGCGGAG 26205
154 ....PheGluGlnAspAlaSerGlyAspGluArgSerGlyProGluAsp 168
||:|||||:
26206 CACCGTCTCGCGCGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 26255
169 SerAsnAspSerArgArgGlySerGly.....SerProGluAspThrPr 183
::|:|||||:
26256 CTGGCGAAGGCTCGGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAG 26305
183 OArgSerSerGlyAlaSerSerIlePheAspLeuArgAsnLeuAlaAla 200
|||:|||||:
26306 TCGCGGCTCTTCGCGCTCAGC.....TCGAAAGCATCGAGCGCGC 26346
200 spSerLeuLeu.....ProSerLeuGluGluArg 209
|||||:
26347 GCTGCTCTCTCAGTCGAGTGAGCGCTCTGCGCGCTCTGCGCGCGAG 26396
210 AlaAlaProGluAspValAspArgArgAsnGluThrLeuArgArgGlnH 226
|||||:
26397 AGGGCGCCCGCTCC.....TCGACCGGTGAGCGTGGTA 26431
226 sArgProProAlaLeuLeuThrLeuTyrProAlaProAspGluAspGlu 242
::|:|||||:
26432 CAGCCGCGCTCTTGGCTCATGTCTCTCGCGCGCGCTGCGCGCTC 26481
243 Ala.....ValGluArgCysSerArgProGluProArgGluH 256
|||:|||||:
26482 GCTCGCGCTCGAGCGCGCGCTGCTGCGCGCACAGCGGCGAGATCG 26531
256 spHeGlyGlnArgIle.LeuValIysCysLeuSerLeuLysPheGluIle 272
::|:|||||:
26532 C.....CGCGCGCTTCGTCGAGGCGGTCTCTCTCTCGAGAGCGGCG 26575
273 GluIleGluProIle.....Ph 278
::|:|||||:
26576 CGCATCGCGCGCTGCGCGAGAAAGCGCTCACACCGTGGCGGACAGCG 26625
278 eGlyIleLeuAlaLeuTyrAspValArgGluLysLysIleSerGluA 295
|||:|||||:
26626 CGGCATGGCGCGCGTC..... 26641
```

295 snPheTyRPhAspLeuAsnSerAspSerMetLysGluLeuAlaArgAla 311
:::|||||:.....:|||||
26642GAGCTGGCGCTCCGACTCCAGACTACCTCGCTCC 26680
312 HisGluThrHisProAlaIleSerThrLeuAla.....Argse 324
|||:::|||||:|||||
26681 TGGGGCGACAGG.....CTCTCCACCGCGCGGTCAACAGCCCGAGGG 26724
324 rAlaIlePheSerValThrTyRProSerProAspIlePheLeuValIleL 341
:|||||:|||||:|||||
26725 TACCTCGTATCCGGCGAGCCCGCGCTCGACGCGCTCTC..... 26767
341 yslLeuGluysValLeuGlnGlnGlyAspIleSerGluCysGluPro 357
|||:::|||||:|||||
26768GACGTCTCCACCGCCACCAAGGTG 26791
358 TyrMetValLeuysGluValAspThrAlaLysAsnLysGluLysLeuG 374
:::|||||:|||||:|||||
26792 TTGCGCCGCAAGATCCGCTCGACTACGCTCCCACTCCGCGCCAGATGA 26841
374 uLysLeuArg.....LeuAlaIaGlnPheCysThrArgLeuGlyA 389
:|||||:|||||:|||||
26842 CGCGGTCCAGAGAGCTCGCGCA..... 26866
389 rGlyThrArgMetProPheAlaTrpThrAlaValHisLeuAlaAsnIleVal 405
|||||:|||||:|||||
26867GGTCTAGCCCAATCGCT 26884
406 SerSerAlaGlyGlnLeuAspArgAspSerAspSerGluGlyLysArg 422
:::|||||:|||||:|||||
26885 CCTCGGACGCGCGAGCTCCCTTTATTGACCGCTCACCGCGCACAGGCT 26934
422 gProAlaTrpThrAspArgArgArgGlyProGlnAspArgAlaSer 439
:::|||||:|||||:|||||
26935 CGAC.....GGCTCG 26945
439 erGlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThr 455
:::|||||:|||||:|||||
26946 AGCTCGAGCGCGCTACTGATCGAAACCTCGCGCAACCGCTCTG... 26992
456 ValThrAsnPhelysGlnGluAlaGluArgLeuSerAspGluAspLe 472
|||:|||||:|||||:|||||
26993TTCTCGAGCGCGACCGAGCGGCTCTCGAC...GATGG 27027
472 uPheLysPheLeuAlaAspMetArg...ArgProSerSerLeuLeuArg. 487
:::|||||:|||||:|||||
27028 GCATCGCTTCTCGGTGAGGTCAAGCCCCCATCCGCTCGACGCTGCGCC 27077
488 ArgLeuArgProValThrAlaGlnLeuLysIleAspIleSerProAlaP 504
|||||:|||||:|||||:|||||
27078 TCCGCGAGACTGCGAGCGCTCACCGCTCGATCCGCTGCTGCGGCTCC 27127
504 oGluAsnProHisPheCysLeuSerPro.....GluLeuLeuHisIle 518
|||:|||||:|||||:|||||
27128 ATTGAGAGAGAGAGGCACTGCGCGCTGCTCTCTCTGCGCGGA 27177
519 LysProTyRProAspProArgGlyArg..... 527
|||||:|||||:|||||:|||||
27178 GCTCTTACCGAGGCGCTCGCGCTCGACTGGAAGACTTCTTGGCGCCT 27227
528ProThrLysGluIleLeuGluPheProAla 538
|||:|||||:|||||:|||||
27228 ACAGTCCCGCGCAAGGTCTCCCTCCCGCAC.....CTACCCCTTCCAGCA 27271
538 rGluValLysAlaProHisThrSerTyRArgAsnLeu..... 550
|||:|||||:|||||:|||||
27272 GAGCGGTTTGGCTCGAGCTCTCCAGGAGAGGCTTCCGAGCTGCGCT 27321
551 LeuTyRValTyRProHisSerLeuAsnPheserSerArgGln..Glyser 566
:::|||||:|||||:|||||:|||||
27322 CCGGAGGCGCTGAGCTGCGCGAGCAATCCGCGTGTGCGGCGCGCTCG 27371

567 ValArgAsnLeuAlaValArgValGlnTyrMetThrGluLysAlaProse 583
|||||:|||||:|||||
27372 CCTTGGCCGACCGCGGTGGCTTTCTT...TTACAGGCGGCTCTCCCTC 27418
583 rGln.....AlaLeuPro..... 587
|||||:|||||:|||||
27419 GCAGAGACACCGCTGGCTCGAAGGCCCATCGCTTCTGCGACACCCATCT 27468
588ValIlePheGluLysSerSerCysSerGluPheThrArgGlu 601
|||||:|||||:|||||:|||||
27469 ACCGGGCAACCGGCTTCTTCGACCTCGCCCTCGACGTCCGCCACCGCGTGC 27518
602 AlaPheThrProValValTyRHisAsnLysSerProGluPheTyRGlul 618
|||:|||||:|||||:|||||
27519 GCCTCGACCGCTC.....GAAGA 27537
618 uPheLysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuP 635
|||:|||||:|||||:|||||
27538 GCTTACGCTCGAGGCGCTCTGCTCTCCATCGAGAGACACCGCTCTC. 27586
635 heThrPheThrHisValSerCysGlnProArgProGlyThrAlaLeuGlu 651
|||:|||||:|||||:|||||
27587CTCCAG 27592
652 ThrProValGlyPheThrTrpIleProLeuLeuGlnHisGlyArg..... 666
|||||:|||||:|||||:|||||
27593 ATCTCGCTGGG.....CCGCTGACGACGACGACGACGAAGGCG 27630
667LeuArgThrGlyProPhe. 672
|||||:|||||:|||||:|||||
27631 GCTCTTCTTCCATAGCCGACAGAGAGAGCGGCTTCAAGATGGCCCTTGA 27680
673CysLeuProVal 676
|||:|||||:|||||:|||||
27681 CTGCGACGSCAGCGCTCTCTGCGCGGCGACCCCATCCCTCTCCGCC 27730
677 SerValAspGlnProProProSerTyRSerValLeuThrProAspValAl 693
:::|||||:|||||:|||||
27731 GATCTCCAGAGTGGCTCTCCCTCGAGTGGCANC.....CCGCTGA 27771
693 aLeuProGlyMetArgTrpValAspGlyHisLysGlyVal...PheSerV 709
|||:|||||:|||||:|||||
27772 CCTCGAAGGCTCTTACGCAACCTCGCCACCTCGGCTTGGCTTACGGCC 27821
709 aGluLeuThrAlaValSerSerValHisProGlnAspProTyRLeuAsp 725
|||:|||||:|||||:|||||
27822 CCGAGTTCAGGCGCTCGCTCGCTTACAGCGCGGAGAGAGAGCTTTT 27871
726 LysPhePheThrLeuValHisValLeuGluGluGlyAlaPheProPheAr 742
|||:|||||:|||||:|||||
27872 GCCGAAGCCMAAGCTCCGGAACGGCGCGAAAGATGCGCGCGGTTTGC 27921
742 gLeuLys.....AspThrValLeuSerGluGlyAsnValGluG 755
|||:|||||:|||||:|||||
27922 CTTCCACCTCGCGCTGCTGACAGCGCCCTGCATGCATGACCTGGCTTTGAG 27971
755 GlnLeuLeuArgAlaSerLeuAla.....Ala 763
|||:|||||:|||||:|||||
27972 ACGAGAGAGAGAGAGAGGTGCTGCTGCGCTTCTGAGAGGAGTCTCG 28021
764 LeuArgLeuAlaSerProGluProLeuValAlaPheSerHisValLe 780
|||||:|||||:|||||:|||||
28022 CTGCGCTCGGTGCGCACACCTTGGCGCGGCTTCCAC..... 28063
780 uAspLysLeuValArgLeuValIleArgProIleThrIleSerGlyGlnI 797
|||:|||||:|||||:|||||
28064CGTCCCAAGGCTGAATCTCCGCT 28088
797 LeValAsnLeuGluArgGlyAlaPheGluAlaMetAlaHisValAlaSer 813
:::|||||:|||||:|||||:|||||
28089 CGATGCTCTGCGCGAGCGCGAGGAGACCTTGTGCTCGGTGCAACG 28138
814 LeuValHisArgSerLeuGluAlaIaGln.....AspAlaArgG 827

```
28139 CTCCGCGATGCGGAGACGCTCCGCGCGACGCTCCGACCCCGGACGCTTC 28188
      |||::: |||::: ::||| |||::: |||:::
827 YHISGYSRPRGGLNLEUAlAATyValNHStyAlAphArLeuProG 844
      |||::: |||::: |||::: |||::: |||:::
28189 CCAACGATGAGCGCTC.....TTCGCGCTGCACT 28217
      |||::: |||::: |||::: |||::: |||:::
844 LyrHgluProSerLeuProAspGlyAlAProProValThrValGlnAla 860
      |||::: |||::: |||::: |||::: |||:::
28218 GGACGCGAGCTCCAAAGCCCACT...TCACGCGCTCGCGCGCGAGCGCG 28264
      |||::: |||::: |||::: |||::: |||:::
861 AlArHrLeuAlAArgGlySer.....G1 868
      |||::: |||::: |||::: |||::: |||:::
28265 GTCCCTTCGCGACGCGGCGCAAGATCTCGCGCTCGACGCCCGCTCGC 28314
      |||::: |||::: |||::: |||::: |||:::
868 YArgProAlAserLeuTyrLeuAlAArgSerLys...SerLeSerSers 884
      |||::: |||::: |||::: |||::: |||:::
28315 CCGCTACGCGGACCTCGCTCGCTCCGACAGCGCCCTCGACAGGGCGCTT 28364
      |||::: |||::: |||::: |||::: |||:::
884 eRAsnProAspLeuAlAAlAAlAProGlySerValAspAspGluValSer 900
      |||::: |||::: |||::: |||::: |||:::
28365 CGCCTCCGCGCTCGCTCGCTCGCC...TTCATGATCGACCGCGAGCG 28411
      |||::: |||::: |||::: |||::: |||:::
901 ArgTLeuAlAserLysLeuLeuNHISgluGluLeuAlAAlAAlAAlA 917
      |||::: |||::: |||::: |||::: |||:::
28412 GACCTCGTCCCGAGC.....GCCACAGC..... 28435
      |||::: |||::: |||::: |||::: |||:::
917 lValSerSerAlAValAlArgGluAlAlAAlAAlAAlAAlAAlAAlA 934
      |||::: |||::: |||::: |||::: |||:::
28436 .....GCCACCGCGCTCGCACTCGCTCTTGCAA...GCCCTGG.... 28471
      |||::: |||::: |||::: |||::: |||:::
934 hePheGluLeuMetValLysSerMetAlAAlAAlAAlAAlAAlAAlA 950
      |||::: |||::: |||::: |||::: |||:::
28472 .....CTCGCGCGAGCAA 28483
      |||::: |||::: |||::: |||::: |||:::
951 ArgLeuAspHrProArgLysLeuArgPheProGlyArgPheLeuAspAs 967
      |||::: |||::: |||::: |||::: |||:::
28484 CGCTCGCTCGTGGCGC..... 28501
      |||::: |||::: |||::: |||::: |||:::
967 rIleThrAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 984
      |||::: |||::: |||::: |||::: |||:::
28502 .....CTGCTCGCTCGCAACCGACGCGCGCTCGCACCCACACCG 28541
      |||::: |||::: |||::: |||::: |||:::
984 lAsLysAspValGluLeuAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1000
      |||::: |||::: |||::: |||::: |||:::
28542 AAGCAGCAGTCAAG..... 28555
      |||::: |||::: |||::: |||::: |||:::
1001 LeuSerAspLeuLeuSerLeuValAspArgGlyPheValPheSerLeuVal 1017
      |||::: |||::: |||::: |||::: |||:::
28556 .....GACCTCGC 28563
      |||::: |||::: |||::: |||::: |||:::
1017 lArgAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1033
      |||::: |||::: |||::: |||::: |||:::
28564 TCACGCGCGCGCTCGGGGCGCTCGCGCGCTCGCGCAAGTGAAGACCCAG 28613
      |||::: |||::: |||::: |||::: |||:::
1033 nProAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1050
      |||::: |||::: |||::: |||::: |||:::
28614 ACCTCGCGCTCT.....C 28627
      |||::: |||::: |||::: |||::: |||:::
1050 eRHSgluNHIStyrValThrLeuAsnLeuProCys..... 1061
      |||::: |||::: |||::: |||::: |||:::
28628 CTGCTGCACTGCACTCGACGAGGCTCCACAGCGCCCTGCTAGCGCC 28677
      |||::: |||::: |||::: |||::: |||:::
1062 .....CysProLeuSerProPro.....AlAserProse 1071
      |||::: |||::: |||::: |||::: |||:::
28678 GCTCGACACAGAGAACGCTCGCGCTCCGCAACGGGAAACCCCTTA 28727
      |||::: |||::: |||::: |||::: |||:::
1071 rProSerValSerSerThrThr..SerGlnSerSerThrPheSerSerGln 1087
      |||::: |||::: |||::: |||::: |||:::
28728 TCCGAGAGTGGCGCAACGCTCGACGAGCGCGCTCATCCGCGCGCAA 28777
      |||::: |||::: |||::: |||::: |||:::
1088 AlAProAspProLysValAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1104
      |||::: |||::: |||::: |||::: |||:::
28778 GCACCCAGCTGGCGCTC..... 28795
      |||::: |||::: |||::: |||::: |||:::
1104 gGlnGlnNHISpHeLeuAlAglYLeuLeuLeuHrGluLeuAlAAlAAlA 1121
      |||::: |||::: |||::: |||::: |||:::
28796 .....CATATTCCGACCAAAAGCACTTCGACGCGCTCGCGCTCGTCG 28838
      |||::: |||::: |||::: |||::: |||:::
1121 euGluProGluAlAglYAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1137
      |||::: |||::: |||::: |||::: |||:::
28839 ACGCCCGCGAGCGCCAG...GGCGCGCTCGGACAGCGCGCAAGTCCGATC 28885
      |||::: |||::: |||::: |||::: |||:::
1138 AlAValNHISerLeu..... 1142
      |||::: |||::: |||::: |||::: |||:::
28886 GCCGTGCAGCGCGCAGGCGCTCAATTCCGATGCTCGACACCCCTTGG 28935
      |||::: |||::: |||::: |||::: |||:::
1143 .....LeuCysAlYHis..... 1146
      |||::: |||::: |||::: |||::: |||:::
28936 CATGTATCCGGCGAGCGCGCGCGCTCGGAGCGCAAGCGCGGCGCATCG 28985
      |||::: |||::: |||::: |||::: |||:::
1147 .....AspThrAspProArgTyrAlAAlAAlAAlAAlAAlAAlAAlA 1161
      |||::: |||::: |||::: |||::: |||:::
28986 TTACTGAAGTGCCTCCAGATGTCTCCGATACACCGTACGCGCGGCGTG 29035
      |||::: |||::: |||::: |||::: |||:::
1162 .....AlaG1 1163
      |||::: |||::: |||::: |||::: |||:::
29036 ATGGGGGCTTCGCGCGCAGCCCTTGCTCCAGCGCATCGCGCGCGCG 29085
      |||::: |||::: |||::: |||::: |||:::
1163 uLeuTyrLeuProLeuLeuSerLleAlAArgAspHrLeuProArgLeuH 1180
      |||::: |||::: |||::: |||::: |||:::
29086 CATGATCTGCGCC.....ATCCCGCAGCGCT 29111
      |||::: |||::: |||::: |||::: |||:::
1180 lAspPheAlAglYlY..ProGlyGlnArgSerArgLeuAlA...SerMe 1195
      |||::: |||::: |||::: |||::: |||:::
29112 GTCCTTCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTAC 29161
      |||::: |||::: |||::: |||::: |||:::
1195 tLeuAspSerAspHrGluGluGluGlyAspAlAAlAAlAAlAAlAAlA 1212
      |||::: |||::: |||::: |||::: |||:::
29162 TATGAGCTGTGATCTCGGC.....ATCGAAACCAATCAACG 29202
      |||::: |||::: |||::: |||::: |||:::
1212 roSerValAlAmet.....AlAlAAlAglYlYProLeuAlAPro 1225
      |||::: |||::: |||::: |||::: |||:::
29203 TGTCTCATTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACG 29252
      |||::: |||::: |||::: |||::: |||:::
1226 GlySerArgAlAserLleSerGlnGlyProProHrAlAAlAAlAAlA 1242
      |||::: |||::: |||::: |||::: |||:::
29253 TCGCAGCGCACCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 29302
      |||::: |||::: |||::: |||::: |||:::
1242 YCysAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1256
      |||::: |||::: |||::: |||::: |||:::
29303 TGGAGCGCTCTCGCGCGCGCTCGGCTTGAGCATGCGCACCTCGGCTCTC 29352
      |||::: |||::: |||::: |||::: |||:::
1257 ..ValLeuTrpValLeuLysAsnThrGlu.....ProAlAAlAAlA 1270
      |||::: |||::: |||::: |||::: |||:::
29353 ACGTACCTGGGCTTCGAGCAGCACTTCCTCGGCTCCAGCATGGGCGCG 29402
      |||::: |||::: |||::: |||::: |||:::
1271 ArgTyrAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlAAlA 1278
      |||::: |||::: |||::: |||::: |||:::
29403 GCATGATGCTGCTCGCTCGACTGTC 29426
      |||::: |||::: |||::: |||::: |||:::
seq_name: /cgn2_6/ptoddata/2/lna/5A_COMB.seq:US-08-588-985-1
seq_documentation_block:
; Sequence 1, Application US/08588985
; Patent No. 5777094
; GENERAL INFORMATION:
; APPLICANT: Michiyuki MATSUDA et al.
; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
```

```
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,985
FILING DATE: January 19, 1996
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 6519 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: spleen cell of homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 24..5519
US-08-588-985-1

alignment_scores:
Quality: 220.00 Length: 2105
Ratio: 0.260 Gaps: 93
Percent Similarity: 40.238 Percent Identity: 18.432

alignment_block:
US-09-736-968a-2 x US-08-588-985-1 ..

Align seg 1/1 to: US-08-588-985-1 from: 1 to: 6519

78 PheProAlaAspAspLeuGluLeuLeuGlnProArgGluCysArgTh 94
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 TTTCCGCTTCATATATTCATCTT.....AAAGAAAGCGATGCT 238
94 rThrGluProGly.....IleProLysAspGluLysLeuA 106
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 TGAAGGAAAGGCGACATGAACAGTATCCCGGCTGACCTCCCTCA 288
106 sPaLaGlnValArgAlaAlaVal.....Glu 114
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
289 TCCAGGAAGTCACACGACACCTCCGAGAGTGTCCACATCTGAGGCG 338
115 MetTyrIleGluAsp..... 119
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
339 CTCTAGCTGCACAGATACAGAGGAGATGTTCCAGAGTCGCGCACATG 388
120 TrpValIleValHisArgArgTyrGlnTyrLeuSerAlaAlaTyrSer 136
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
389 CTATGACCTTATGTAATGCGATCACAATCTTCTGGAAGT..... 431
136 roValThrThrAspThrGlnArgGluArgGlnLysGlyLeuPro..... 150
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
432 ..CTGCGCTAGAGTGAAGTCAAGAAAGTGAAGAAAGTCAAGCCAAA 479
151 .....ArgGlnValPheGluGlnAspAlaSerGlyAspGluArg 163
```

```
480 ATTGATTAATGGAACAGAAATCTAGATTGTGACCTGGTGGTTAGAGT... 527
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
163 gSerGlyProGluAspSerAsnAspSerArgArgGlySerProG 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
528 .....GAGATGGGATATTTG..... 545
180 LuAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuArg 196
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
546 .....GATCCAGAAATTAATGACAGATTAGTCTCTCAGAGCTCATGAA 590
197 LeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnLysArgAlaAlaProG 213
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
591 ATAGCTTCT.....AAACAAGTGGAGGAAGGTTTACAGAGGA 628
213 LuAspValAspArgArgAsnGluThrLeuArgArgGlnHisArgProPro 230
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
629 AAATCTCAAAAGCAGAACATAGATATTACAGCAAGCCAAAGTTTCTG 678
230 LaLeu.....LeuThrLeuTyrProAlaProAspGluAspGluAlaVal 244
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
679 CAACCCCTCTCTGCGCTTGT..... 701
245 GluArgCysSerArgProGluProProArgGlnHisPheGlyLysArg 261
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
702 .....GT 703
261 eLeuValLysCysLeuSerLeuLysPheGluIleGluIleGluProIle 278
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
704 GACCTCAAAATGATGCTTGTAAATAGAGAGAGATGCTGAATC... 749
278 heGlyIleLeuAlaLeuTyrAspValArgGluLysLysLysIleSerGlu 294
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
750 ....CTCATGCTCTATATGATGACCTGAGTCCAAATTCATCAGAG 794
295 AsnPheTyrPheAspLeuAsnSerAspSerMet...LysGlyLeuLeuArg 310
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
795 AACTACCTGCTTCGCTGCTCCAGTTCAAGATTACTTAAGACATAGACAG 844
310 GaLHisGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIle 327
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
845 ATTACATATTTGCGAGCGCTTACTGACCTCGA..... 881
327 heSerValThrTyrProSerProAspIlePheLeuValIleLysLeuGlu 343
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
882 .....AGC 884
344 LysValLeuGlnGlnGlyAspIleSerGluCysCysGluProTyrMetVal 360
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
885 AAAGACCTGAAGAGGAGAAATCAGTTTGTCTGTCAG..... 923
360 IleuLysGluValAspThrAlaLysAsnLysGluLysLeuAla 377
923 ..... 923
377 rGluAlaAlaGluGlnPheCysThrArgLeuGlyArgTyrArgMetPro 393
924 .....ATGTGCGTGGGTCCAGTGAAGCTG... 950
394 PheAlaTrpThrAlaValHisLeuAlaAsnIleValSerSerAlaGly 410
950 ..... 950
410 nLeuAspArgAspSerAsp.....SerGluGlyLysArgArg 423
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 .....AGGCAACAACACACAGAAACTGACCTCGGGTTCGCGGAC 993
423 roAlaTrpThrAspArgArgArgArgGlyProGlnAspArgAlaSerSer 439
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
994 CT..... 995
440 GlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThrVal 456
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

996 .....TTTGAGTGGCTGTGATGATGCT 1018
456 1ThrAspPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLeu 473
      |||.....:
1019 AACAGATATATAAT.....GGAAGAGTAGATGATGAAGATTAAC 1059
473 heLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArgArgLeu 489
      |||.....:
1060 ACCATTTCATTTCC.....TTT 1076
490 ArgProValThrAlaGlnLeuLysIle.....AspIle 500
      |||.....:
1077 CAGCCGGGTGGCAGGAGAAATGACTTCCTTCACAGCTGTTATAACAAGCT 1126
500 eSerProAlaProGluAsnProHis.....P 509
      |||.....:
1127 CATCGCTGCCAAGAAGATCAACCAAGGGCGAGGTTGTGTGGTAACT 1176
509 heCysLeuSerProGluLeuLeuHis.....IleLysProTyrProAsp 523
      |||.....:
1177 TGAATTACTTCCTGGAGATATCCATCATCCGAAAGAGTTTCCGCAT 1226
524 ProArgGlyArgProThrLysGlnIleLeuGluPheProAlaArgGluVal 540
      |||.....:
1227 TTAGTGCACAGGACACAGCTGTGCTCGAAAAACAGGGTTTCCGAGAT 1276
540 1TyraIaProHisThrSerTyrArgAsnLeuLeuTyrValTyrProHis 557
      |||.....:
1277 AATCATGCGCT...GCTGATGTTGAAATGATATCTATGTACATTAAGTTC 1323
557 eLysLeuAsnPhe...SerSerArgGlnGlySerValArgAsnLeuAlaVal 572
      |||.....:
1324 AAGGAGATTTTGTATTAAGAGCAAAACAACAGCAAGAACTGGAGTGC 1373
573 ArgValGlnTyrMetThrGlnGluAspProSerGlnAlaLeuProValIle 589
      |||.....:
1374 ACGGTGTCT...GTGTACAGATGAGGTGGAAACGATTAGACAGTGTGAT 1420
589 ePhe.....GlyLysSerSerCysSerGluPheThrArgGluAlaIle 603
      |||.....:
1421 TTTCCGGGGTGGTGGTGAAGCGATTTCAGAGTCAAA..... 1460
603 heThrProValIleValTyrHisAsnLysSerProGluPheTyrGlnGluPhe 619
      |||.....:
1461 ..TCTGTGATTACTACCAAGTAAGCACAGCCAGCTGTGAGACTGTT 1508
620 LysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuPheThr 636
      |||.....:
1509 AAGGTGGCCATTCCTCATCGAGAGCGTTAAACCGCAGTCACCTTCGGTTTAC 1558
636 rPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGluThrP 653
      |||.....:
1559 CTTCGCCACAGGTCTACAGAGACTCTTAAGATTAATCT...GAGAAA 1605
653 roValGlyPheThrTrpIleProLeuLeuGlnHisGly.....ArgLeu 667
      |||.....:
1606 TATTGGCACTACATTTGTCAAGCTGATGATAGATGTTACCAACCTTG 1655
668 ArgThrGlyProPheCysLeuProValSerValAspGlnProProPhe 684
      |||.....:
1656 CAGAGCGGAGAGACAGCATTTATCTCTATAAGCCGGAAGCGAAGACGCT 1705
684 rTyrSerValLeuThrProAspValAlaLeuProGlyMetArg..... 698
      |||.....:
1706 GGAAGATGCTGCACG...TACTTGAAGTGTCCCTCCACGAAGGACAGAT 1752
699 ..TyrValAspGlnHisLysGlyValAlaPheSerValGlnLeuThrAlaVal 714
      |||.....:
1753 TGGAAAGAAAGGGCCATCGCCACCGCAAGACGATGACAGACCTTGGG 1802
715 SerSerValHisProGlnAspProTyrLeuAspLysPhePheThrLeuVal 731
      |||.....:
1803 ACGTCGACCATTTAGCAAGAGACTCTTC.....CAGATCTCCACCGCTCT 1846

```

```

731 1.....HisValLeuGlnGluGlyAlaPhePro 741
      |||.....:
1847 GTGCTCCACCAACTGACTGAGAACGTGGACCTTGTGGGCTCTTGAAT 1886
741 heArgLeuLysAspThrValLeuSerGlnGluAsnValGlnGlnGluLeu 757
      |||.....:
1897 GCGCTCCACACACGCTGCTGCAGAG..... 1925
758 ArgAlaSerLeuAlaAlaLeuArgLeuAlaSerProGluProLeuValAl 774
      |||.....:
1926 .....AACTGAGCGCAGCTGATGAAGTGCATGGTGTGTAAGTAGTGA 1969
774 aPheSerHisValLeuAspLysLeuValArgLeuValIleArgPro 791
      |||.....:
1970 GTTCTTTCAGACACAGCTTGATGCCCTCTTCACATCATGATGAG... 2015
791 rolIleLeSerGlyGlnIleValAsnLeuGluArgGlyAlaIlePheGluAla 807
      |||.....:
2016 .....AACTCAGAGAGTGAGACTTTTGACAGCTTAGCTTGTGATGCT 2057
808 MetAlaHisValAlaSerLeuValHisArgSerLeuGlnAlaAlaGlnAs 824
      |||.....:
2058 CTGTATTTCATCATTCGACTGATT.....CCTGATAGAAA 2092
824 pAlaArgGlyHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPhe 841
      |||.....:
2093 ATTTCAGCATTTTAATCTGTTTGGAACTTACATTACAGAACACTTT. 2141
841 rGluProGlyThrGluProSerLeuProAspGlyAlaProProValThr 857
      |||.....:
2141 ..... 2141
858 ValGlnAlaIleThrLeuAlaArgGlySerGlyArgProAlaSerLeuTyr 874
      |||.....:
2142 .....AGTGCACGTTAGCC.....TA 2158
874 rLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAlaVal 891
      |||.....:
2159 CACGAGTGTGACAAA...GTGTTGAAGAACTACGTGACGGTGTCTGACA 2205
891 lApProGlySerValAspAspGluVal..... 899
      |||.....:
2206 AGCCGGGA...GTAAATGACAGCTGTACAAAGCCATGAAGCGCTGAAA 2252
900 .....SerArgIleLeuAlaSerLysLeu 908
      |||.....:
2253 TCCATCTTCAAGTTCATCGTGGCTCCAGATCTGTTCATTAACACTGTA 2302
908 uHisGlnGluLeuAlaLeuGlnTrpValIleSerSerSerAlaValArg 925
      |||.....:
2303 TGAAGACAAAGGAGAGGCTGACTCGTG.....G 2331
925 lValAlaIleLeuGlnHisAlaTrpPhePheGlnLeu.....MetVal 939
      |||.....:
2332 AATTTTGTGCTGCAG.....CTTTCAGGTCCATCAATGCAATGATGAT 2372
940 LysSerMetAlaLeuHisLeuLeuGlnGlyAlaArgLeuAspThrProAr 956
      |||.....:
2373 AGCAGCATGTCA..... 2384
956 gLysLeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValG 973
      |||.....:
2385 .....GACACAGACCGTCCGGGTGA 2403
973 LysSerValGlyLeuGluValIleThrArgValHisLysAspValGluLeu 989
      |||.....:
2404 AGGGGCGACAGCTGAATACTTACCAACGATCGTCAACGATGTGAATTG 2453
990 AlaGlnHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeu 1006
      |||.....:
2453 ..... 2453

```

```
1006 rLeuValAspArgGlyPheSerLeuValArgAlaHisTyrLysG 1023
2454 .....CTGTTT
1023 lValAlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuThr 1039
2460 .....GATCCAAAGAGCTCAGCAAA 2480
1040 LeuArgMetGluPheThrArgAlaLeuGlySerHisGlnHisTyrValTh 1056
2481 ATGTTTACTGAATTC.....AT 2497
1056 rLeuAsnLeuProCysCysProLeuSerProProAlaSerProSerPro 1073
2498 CCTCAATGTTCC..... 2510
1073 eValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaPro 1089
2510 ..... 2510
1090 AspProLysValThrSerMetPheGluLeuSerGlyProPheArgGlnG 1106
2510 ..... 2510
1106 nHisPheLeuAlaGlyLeuLeuLeuThrGluLeuAlaLeuAlaLeuGlu 1123
2511 .....ATGGCTTCTGACCATCCAGAACTCTACTGCTTG.... 2546
1123 roGluAlaGluGlyAlaPheLeuLeuHisLysLysAlaIleSerAlaVal 1139
2547 .....ATCGAAATCGTC 2558
1140 HisSer...LeuLeuGlyHisAspThrAspProArgTyrAlaGluAl 1155
2559 CACAGTACCTCTTCACACAGCATGAC..... 2585
1155 aThrValLysAlaArgValAlaGluLeuTyrLeuProLeuLeuSerIleA 1172
2586 .....TGCAGAGAGATCCCTGCCATGATGAC.... 2614
1172 lAargAspThrLeuProArgLeuHisAspPheAlaGluGlyProGlyGln 1188
2614 ..... 2614
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThrGluGlyGlyAs 1205
2615 .....CGATCAGCTCAAGTACCATCTGGAGAGACAGAGAGACCTGGAGG 2658
1205 pIleAlaGlyThrIleAsnProSerValAlaMetAlaIleAlaGlyGlyP 1222
2659 CCTGCTGTCAGCTGCTCAGCCACAT.....CCTGGAGGTG 2693
1222 roLeuAlaProGlySerArgAlaSerIleSerGlnGlyPro..... 1235
2694 CTGTACAGAAAGACGTGGGGCCAAAGAGAGAGAGCTCAGATTTATCAT 2743
1236 .....Pro.ThrAlaSerArgAlaGlyCysAlaLeuSerAlaG 1248
2744 GGAGAAACTTCTCCGGACCGTGAACCGAACCAGTATTCATGAGAGAG 2793
1248 luserSer.....ArgThrLeuLeuAlaCysValLeuTrpValLeuLys 1262
2794 ATTGTGAACATATTGGAAACTCTGTGCTTCATGACAGCTATTATTACGA 2843
1263 AsnThrGluProAlaLeuLeuGlnArgTrpAlaThrAspLeuThrLeuPr 1279
2844 CAATATGGA..... 2852
1279 ogInLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheG 1296
2853 .....GATTACCATTTATGCCCACTTGATCAG.... 2879
1296 luyrLysGlyLysLysAlaPheGluArgIleAsnSerLeuThrPheLys 1312
2880 .....ACTTTGGG 2888
1313 lysSerLeuAspMetLysAlaArgLeuGluGluAlaIleLeuGlyThrI 1329
2889 AAA.....ATGAGACATGATGTGGAGATTTCCTAATGAAGAACTT 2929
1329 eGlyAlaArgGlnGluMetValArgArgSerArgGluArgSerProPhe. 1345
2930 CATCATGTTTAAAGAACTCATT.....GGAAGAACCTTTACCCCTTCG 2973
1346 .....GlyAsnProGluAsnValArgTrpArgLysSerVal 1357
2974 ACTGGGTATCATGAACATGTCGCAAAATTAAGTTCTTCGAGACAAAT 3023
1358 ThrHisTyrLysGlnThrSerAspArgValAspLysThrLysAspGlu 1374
3024 AATCAGTATGACAGATATGCTGAAACAA.....AATTTCTGATCAGAC 3067
1374 tGluHisGluAlaLeuValGluGlyAsnLeuAlaThrGluAlaSerLeuY 1391
3068 CAACCTTGAGCTACAGCTGTGGAAACACTACTTTCACCTGCT.....G 3111
1391 aValLeuAspThrLeuGluIleIleValGlnThrValMetLeuSerGlu 1407
3112 TTGCTTTCTTACTCAAGATCCCTGCAAGTGGAGAAATTTTCAAGTCC 3161
1408 AlaArgGluSerValLeu.....G 1414
3162 AAGAGAGCCAAATCCTTAAACAATGAGATATGAGAGACAGATGG 3211
1414 yAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAla 1431
3212 CTTTGAATCAGACAGATGTGGTCAACCTTGT..... 3245
1431 euPheLeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhe 1447
3246 .....CAACCAAGATTAAGTTCATTCGCAAAATGTGGCCCAATA 3287
1448 ProGluLeuLeuPheGluGluAspThrGluLeuCysAlaAspLeuCysLe 1464
3288 TTAGAATGACATTAAATCCGAGAGAGGAGCTG..... 3320
1464 uArgLeuLeuArgHisCysGlySerArgIleSerThrIleArgThrHisA 1481
3321 .....CGCAAGCCACCATC..... 3335
1481 lAserAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGluIleGlyHis 1497
3336 .....CCCATCTTCTTTGATATGATGACAGTGTGAATTCATTCGACCGCA 3380
1498 AsnPheAlaArgValLysMetGlnValThrMetSerLeuSerSerLeuVa 1514
3381 ACTTCCAAATGTTTGAATAATGATCATACCAAGCTGATCATGAGAAT 3430
1514 l.....GlyThrThrGln.....AsnPhe 1521
3431 CGAAGAGGAGAGAGAGAGACAGCAACAGTATATTGATAAATCC 3480
1521 eGluGluHisLeuArgArgSerLeuLysThrIleLeuThrTyrAlaGlu 1537
3481 TTCTGGAACACTCGCAGAG.....CAC 3503
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGluGlnValGlnAspLe 1554
3504 AAATACCTCGCAAAACAGAGAAACTTTTGAAACGCTGTGCTGCGCTT 3553
1554 uMetPheAsnLeuHisMetIleLeuThrAspThrValLysMetLysGluH 1571
3554 AATGAAAGCTTTTGATTTATGAACCATCATG.....C 3588
1571 lSGlnGluAspProGluMetLeuIleAspLeuMetTyrArgIleAlaArg 1587
```

3589 ACGACGAGACAAAGAAAACCGCATGAGCTGACCGTCAATGTGCTGAAAT 3638
1588 GLYTYRGLN.....GlySerProAspLeuArgLeuThrTrpLeuGlnAs 1602
3639 TTCTACAAAGAAATTTGAAGAGAAATGATATATAGGATATTGTGTCAA 3688
1602 nmetAlaGlyLysHisAlaGluLeuGlyAsnHisAlaGluAlaAlaGlnC 1619
3689 GCTCTGTGACCTGCACAGAGGTGTGATTAACCTACACCGAAGCGCTTACA 3738
1619 ysmetValHisAlaAlaAlaLeuValAlaGluTrpLeuAlaLeuLeuGlu 1635
3739 CTTTGCTT.....CTCCATCGAAAGCTTTTAAGTGGTGGAG 3776
1636 AspGlnArgHisLeuProValIGlyCysValSer..... 1646
3777 GATGTG.....TGTGTGGCCACCTCACCCAGCGGGA 3808
1647PheGlnAsnIleSerSerAsnValLeuGluGluSerAlaIleSerA 1662
3809 CGGGTACACGACCCACAGCAGGAGCTGAAGAGCAGCTGTACAGG 3858
1662 spARPIleLeuSerProAspGluGluGlyPheCysSerGlyLysHisPhe 1678
3859 AAATCATC.....CACTACTTCGACAAAGGCAAGATGTGG 3893
1679 ThrGluLeuGlyLeuValGly.....LeuLeuGluGlnAlaAlaGlyTY 1693
3894 GAGAGGCGCATTTGGCTTGGCGAAGAGACTAGCCGACAG..... 3932
1693 rPheThrMetGlyGlyLeuTrpGluAlaValAsnGluValTYrLysAsnL 1710
3933TATGAG.....AACGAATGTTT..... 3950
1710 euIleProIleLeuGluAlaHisArgAspTYrLysLysLeuAlaAlaVal 1726
3951GATTATGAGCAACTCAGCGAATTTG 3974
1727 HisGlyLysLeuGlnGluAlaPheThrLysIleMetHisGlnSerSerGI 1743
3975 CTGAAGAAACAGCGCTCAGTTTATGAACATCTGTCAAGTGAATC.... 4019
1743 YTrpGluArgValPheGlyThrTYrPheArgValGlyPheTYrGlyAlaH 1760
4020AGGCCAAGCCTGACTATTTGCTGTGGCTACTAGCAGACAG 4062
1760 IspHegLysP...LeuAspGluGlnGluPheValTYrLysGluProSer 1775
4063 GGTTCGCCACACTTCTCGGGGAAAAGTTTCAATTACCGAGGAAAGAG 4112
1776 IleThrLysLeuAlaGluIleSerHisArgLeuGlnGluPheTYrThrGI 1792
4113 TATGAGCCCGCGAAGATTTTGAAGCTCGGCTTAACTCAGTTTCCAA 4162
1792 uArgPheGlyAspAspValValGluIleIleLysAspSerTYrTrpValA 1809
4163 C.....GCCGAGAAATGAAGACAACTCTCCACACAG 4194
1809 sPlySerLysLeuAspSerGlnLysAlaTYrIleGlnIleThrTYrVal 1825
4195 GCGACGATATTAAACTCTCTGCGCAAGTATATTCAGTCTTCACAGTG 4244
1826 GluProTYrPheAspThr.....TYrGluLeuLysAspAR 1837
4245 AAGCCCAAACTCGATCGCCTTAAGTTTACAGCGCCACTGTCCAGAGCA 4294
1837 gValThrTYrPheAspArgAsnTYrGlyLeuArgThrPheLeuPheCysT 1854
4295 GATTGTAAAGTTTACAGGCTGAACGAGGTCGACGATTTGAATATTCCTC 4344
1854 hrProPhe.....ThrProAspGlyArgAlaHisGlyGlu 1865
4345 GGGCAATCCGAGAGGAGAGAAACCCAGAC.....AATGAA 4382

1866 LeuProGluGlnHisLysArgLysTYrLeuLeuLeuSerThrAspHisAlaPh 1882
4383 TTTCGCAATATGTGATCGAGAGAACCATATATACAACTGCTATTAATTT 4432
1882 eProTYrIleLysThrArgIleArgValCysHisArgGluGluThrValL 1899
4433 ACTGGAATTTTAAGGTGTGTGAGGTCAAGTCTGTTTCATGATGGAAA 4482
1899 euThrProValGluValAlaIleGluAspMetGlnLysLysThrArgGlu 1915
4483 TCAGCCCTTGAGATGCCATCGAGACCATGACCTGACGAGACAGCAAG 4532
1916 LeuAlaPheAlaThrGluGlnAspProProAspAlaLysMet..... 1929
4533 ATCAACAGCATGTGTGACGACGACCTGGATGACCCACGCTCCCATCAA 4582
1930LeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGln 1945
4583 CCGCTCTCATGCTCCCTGACGCGATCGTGACCCAGCTGTATGGGG 4632
1945 IyProLeuGluValAlaGlnValPheLeuAlaGlu.....IleProGlu 1959
4633 GCTTCGCAAACTACGAAAGGCTTCTTTACAGACCGGTACCTGCAGAG 4682
1960 AspProLysLeuPheArgHisHisAsnLysLeuArgLysCysPheLysAs 1976
4683 CAOCCTGAGCCCGCATGAAAGATCGAAGAGCTC.....AAGA 4720
1976 pPheCys..... 1978
4721 CCGATTGCTTGGCAGATTCCTTTCTGCGCGAAGGATCAGAAATCATG 4770
1979 ..LysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1994
4771 GAGACAAAGTCAAGGAGCGACTGAG..... 4796
1995 AspGlnLysGluTYrHisArgGluLeuGluLysArgAsnTYrCysArgLeuAr 2011
4797CCGTCACAGAGAGATGAGCGCTTTCAAACAGCTGAA 4837
2011 gGluAlaLeuGln 2015
4838 GGAAGAGGTGGAG 4850
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-971-988-1
seq_documentation_block:
: Sequence 1, Application US/08971988
: Patent No. 5786461
: GENERAL INFORMATION:
: APPLICANT: Michiyuki MATSUDA et al.
: TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wenderoth, Lind & Ponack
: STREET: 805 Fifteenth Street, N.W., #700
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
: COMPUTER: IBM compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/971, 988
: FILING DATE: 17-NOV-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/588, 985
: FILING DATE:

ATTORNEY/AGENT INFORMATION:
 NAME: Warren M. Cheek, Jr.
 REGISTRATION NUMBER: 33,367
 REFERENCE/DOCKET NUMBER:
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-371-8850
 TELEFAX:
 TELEX:
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 6519 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOETHERICAL:
 ANTI-SENSE:
 FRAGMENT TYPE:
 ORIGINAL SOURCE: spleen cell of homo sapiens
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 24..5619
 US-08-971-988-1

alignment_scores:
 quality: 220.00 length: 2105
 ratio: 0.260 gaps: 93
 Percent Similarity: 40.238 Percent Identity: 18.432

alignment_block:
 US-09-736-968A-2 x US-08-971-988-1 ..

Align seg 1/1 to: US-08-971-988-1 from: 1 to: 6519

```

78 pheprolaasparpleugluLeuLeuLeuInProArGluCysArgTh 94
|||||
201 TTTCCTGCTCATATTCATCTT.....AAGAGCGGATAGT 238
94 rThrGluProGly.....IleProLysAspGluLysLeuA 106
|||
239 TGAAGGAAAGGCGACATGAAACAGTCATCCGGGTGACCTCCCCCTCA 288
106 spAlaGluValArgAlaAlaVal.....Glu 114
|||||
289 TCCAGGAGATCACCGACACTCCGAGAGTGTCCACCATCTGGAGCGAG 338
115 MetYrIleGluAsp..... 119
|||||
339 CTCACGTGCAAGATACAGGAGATGTTTGAAGTGTGGCGCACATGAT 388
120 .TrpValIleValHisArgArgTyrglnTyrlLeuSerAlaAlaTySerP 136
|||
389 CTAATGACCTATTATGAGCGATCACAATCTTTCTGGAAGT..... 431
136 roValThrThraspThrGlnArgGluArgGlnLysGlyLeuPro..... 150
|||
432 .CTGCCTCAGATGAACTCAAGAACTGAAGAGAAAGTCAACAGCCAAA 479
151 .....ArgGlnValPheGlnGluInAspAlaSerGlyAspGluAr 163
|||||
480 ATTGATTATGAAACAGAAATTCATGATTGGACCTGCTGTTAGAT... 527
163 gserGlyProGluAspSerAsnAspSerArgArgGlySerGlySerProG 180
|||||
528 .....GAAGATGGGAATATTTG..... 545
180 lAspThrProArgSerSerGlyAlaSerSerIlePheAspLeuArgAsn 196
|||
546 .....GATCCAGAAATTAATCAGACGATTAAGTCTTCAGAGCTCATGAA 590
197 LeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArgAlaAlaProGl 213
|||||

```

```

591 ATAGCTTCT.....AACAACTGGAGAAAGCTTACAAAGAGA 628
213 uAspValAsparArgAsnGluThrLeuArgArgGlnHisArgProProA 230
|||
629 AAAATCTCAAAAGCAGAAATGATATTAACAGAACCAAGATTGCTGT 678
230 lAleu.....LeuThrLeuTyrrProAlaProAspGluAspGluAlaVal 244
|||
679 CAACCCCTTCTCTGGCCTTGT..... 701
245 GluArgCysSerArgProGluProProArgGluHisPheGlyGlnArgI 261
702 .....GT 703
261 eLeuValLysCysLeuSerLeuLysPheGluIleGluIleGluProIleP 278
|||
704 GAACCTCAAAAGTGTGTTGTAATAATAGAGAAATGCTGAAGTC... 749
278 heGlyIleLeuAlaLeuTyrrAspValArgGluLysLysLysIleSerGlu 294
|||||
750 .....CTCATGTCTATATGACCTGTGGAGTCCAAATTCATCATGAG 794
295 AsnPheTyrrPheAspLeuAsnSerAspSerMet...LysGlyLeuLeuAr 310
|||||
795 AACTACCTGTTGCTGCTGTCGACATTCAGATTAACCTAAAGACATAGACAG 844
310 gAlaHisGlyThrHisProAlaIleSerThrLeuAlaArgSerAlaIleP 327
|||
845 ATTACATAATTTGCGAGCCGCTTTACTGACCTCGGA..... 881
327 heSerValThrTyrrProSerProAspIlePheLeuValIleLysLeuGlu 343
882 .....AGC 884
344 LysValLeuGlnGlnLysPheLysSerGluCysGluProTyrrMetVal 360
|||
885 AAGACCTGAAAGGAGAAATCAGTTTGTCTGTCTGACG..... 923
360 lLeuLysGluValAspThrAlaLysAsnLysGluLysLeuGluLysLeuA 377
923 ..... 923
377 rGluAlaAlaGluGlnPheCysThrArgLeuGluArgTyrrArgMetPro 393
|||
924 .....ATTGTCGCGTGGTGGCATGAGAGCG... 950
394 PheAlaThrThrAlaValHisLeuAlaAsnIleValSerSerAlaGly 410
950 ..... 950
410 nLeuAspArgAspSerAsp.....SerGluGlyGluArgArgP 423
|||
951 .....AGGAGACACACACAGAAATGACCTCGGGGTGTCGCGCAG 993
423 roAlaThrThrAspArgArgArgGlyProGlnAspArgAlaSerSer 439
|||
994 CT..... 995
440 GlyAspAspAlaCysSerPheSerGlyPheArgProAlaThrLeuThryA 456
996 .....TTTGAGTGTGCTGTGATGATGT 1018
456 lThrAsnPhePheLysGlnGluAlaGluArgLeuSerAspGluAspLeuP 473
|||
1019 AACGAGATTAATAAAT.....GGAAGAGTGAAGATGAAGATGAAGC 1059
473 heLysPheLeuAlaAspMetArgArgProSerSerLeuLeuArgGlu 489
|||
1060 AGCATTTTCATTTCC.....TTT 1076
490 ArgProValThrAlaGlnLeuLysIle.....AspI 500
|||
1077 CAGCCGGTGGCAGGAGAAATGACTCTTCAGACGTGTATTAACAAGAGT 1126

```

```

500 eSerProAlaProGluAsnProHis.....P 509
    : ||| ||| |||
1127 CATCGCTGCGCAAGAGTCAACACAGAGGGCTTTGGGTAACT 1176
509 heCysLeuSerProGluLeuHis.....IleYsProTyProAsp 523
    : ||| ||| :||| |||
1177 TGAAATTACTCTGGAGATATCATCGATCCGAAAGAGTTCCGAT 1226
524 ProArgGlyArgProThrLysGluIleLeuGluPheProAlaArgGlu 540
    : ||| ||| ||| |||
1227 TTAGTGGACAGACACAGCTGTGGCTGAAAAACAGGTTTCCGGAGAT 1276
540 I TyrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTyrProHis 557
    : ||| :||| ||| :||| |||
1277 AATCATGCT...GGTATGTTGCAATGATATCTATGTAACATTACTTC 1323
557 erLeuAsnPhe...SerArgGlnGlySerValArgAsnLeuAlaVal 572
    : ||| :||| ||| :||| |||
1324 AAGGAGATTTGATAAAGAAACAAACACAGCGAAGAGTGAGAGTC 1373
573 ArgValGlnTyrMetThrGlyLysProSerGlnAlaLeuProVal 589
    : ||| :||| ||| :||| |||
1374 ACGGTGCT...GTTCAGATGAGATGGAAAGATTAAGACATGAT 1420
589 erPhe.....GlyLysSerCysSerGluPheThrArgGluAla 603
    : ||| ||| :||| ||| :||| |||
1421 TTTCGCGGTCTGTGATGAAGCATTTTCAGATCAAA..... 1460
603 heThrProValTyrHisAsnLysSerProGluPheTyrGluGluPhe 619
    : ||| :||| ||| :||| ||| :||| |||
1461 ..TCTGATTTACTACCAAGTAAGACGCCAGCTGTGTTGAGACTGT 1508
620 LysLeuHisLeuProAlaCysValThrGluAsnHisLeuLeuPhe 636
    : ||| :||| ||| :||| ||| :||| |||
1509 AAGGTGGCCATTCGCATCGAGAGCGTTAACCGCATCTGGTTTAC 1558
636 rPheTyrHisValSerCysGlnProArgProGlyThrAlaLeuGluThr 653
    : ||| ||| ||| ||| |||
1559 CTTCGCCACAGCTCATCACAGACTTAAGATAAATCT...GAGAAAA 1605
653 roValGlyPheThrPheLeuLeuGlnHisGly.....ArgLeu 667
    : ||| :||| ||| :||| ||| :||| |||
1606 TATTTCAGACATTTGTCAACTGATGATGATCATGATGATCCACCTTG 1655
668 ArgThrGlyProPheCysLeuProValSerValAspGlnProTyrPro 684
    : ||| ||| ||| ||| |||
1656 CGAGACGGAGACACGATCTTATCGTCTATAAGCCGAAAGCAAGCT 1705
684 rTyrSerValLeuThrProAspValAlaLeuProGlyMetArg..... 698
    : ||| :||| ||| :||| ||| :||| |||
1706 GGAAGATGCTGCCAG...TACTTGAGTCTGCCCTCCACGAAGCGAGCT 1752
699 ..TyrValAspGlnHisLysGlyValPheSerValGluLeuThrAlaVal 714
    : ||| :||| ||| :||| ||| :||| |||
1753 TGGAAAGAAAGGCCACTCGGCCACGGGAAGCATGACAGACCTTGGG 1802
715 SerSerValHisProGlnAspProTyrLeuAspLysPhePheThrLeu 731
    : ||| :||| ||| :||| ||| :||| |||
1803 AGCTGCACCATTAAGCAAGACTCTCTC.....CAGATCTCACAGCTCT 1846
731 L.....HisValLeuGluGluGlyAlaPhePro 741
    : ||| :||| ||| :||| ||| :||| |||
1847 GTGCTCCACCAACTGACTCAGAACCTGGACCTTGGGGCTTTGAAT 1896
741 heArgLeuLysAspThrValLeuSerGluGlyAsnValGluGluLeu 757
    : ||| :||| ||| :||| ||| :||| |||
1897 GGGGCTCCAAACACAGCTGTGACAGAG..... 1925
758 ArgAlaSerLeuAlaLeuArgLeuAlaSerProGluProLeuValAl 774
    : ||| :||| ||| :||| ||| :||| |||
1926 .....AACTTGAGCGAGCTGATGAAGTCATGCTGTAAGTAGTGAA 1969

```

```

774 aPheSerHisHisValLeuAspLysLeuValArgLeuValIleArgPro 791
    : ||| :||| ||| :||| ||| :||| |||
1970 GTTCTTCAGACACGTTGATGCCCTTCAACATCATGATGAG.... 2015
791 rolIleSerGlyGlnIleValAsnLeuGlyArgGlyAlaPheGluAla 807
    : ||| :||| ||| :||| ||| :||| |||
2016 .....AACTAGACAGTGAACATTTTGACACGTTACTCTTTGTGCT 2057
808 MetAlaHisValValSerLeuValHisArgSerLeuAlaAlaGlnAs 824
    : ||| :||| ||| :||| ||| :||| |||
2058 CTGGATTTATCATGTGACTGANT.....GCTGATGAA 2092
824 pAlaArgGlyHisCysProGlnLeuAlaAlaTyrValHisTyrAlaPhe 841
    : ||| :||| ||| :||| ||| :||| |||
2093 ATTTCAGCATTTTAATCTGTTTGGAACTTACATTAAAGAAACATT. 2141
841 rGluProGlyThrGluProSerLeuProAspGlyAlaPheProValThr 857
    : ||| :||| ||| :||| ||| :||| |||
2141 ..... 2141
858 ValGlnAlaIleThrLeuAlaArgGlySerGlyArgProAlaSerLeu 874
    : ||| :||| ||| :||| ||| :||| |||
2142 .....AGTCAACGTTAGCC.....TA 2158
874 rLeuAlaArgSerLysSerIleSerSerSerAsnProAspLeuAlaVal 891
    : ||| :||| ||| :||| ||| :||| |||
2159 CACGAAGTTGACAAA...GTGTTGAAGAACTACGTGACGGTGTGAGA 2205
891 lArProGlySerValAspAspGluVal..... 899
    : ||| :||| ||| :||| ||| :||| |||
2206 ACCCGGA...GTAATGACAGCTGTACAAAGCCATGAAGCGCTAGA 2252
900 .....SerArgIleLeuAlaSerLysLeu 908
    : ||| :||| ||| :||| ||| :||| |||
2253 TCCATCTTCACTGATATGTCGCTCCAGACCTGTTCAATCAACTGTA 2302
908 uHisGluGluLeuAlaLeuGlnTyrValValSerSerSerAlaValArg 925
    : ||| :||| ||| :||| ||| :||| |||
2303 TGAACAAGGAGAGGCTGCTGCTG.....G 2331
925 lAlaIleLeuGlnHisAlaThrPhePheGlnLeu.....MetVal 939
    : ||| :||| ||| :||| ||| :||| |||
2332 AATCTTCTCTGAG.....CTCTCAGCTCATCAATGATGATG 2372
940 LysSerMetAlaLeuHisLeuLeuGluGlnArgLeuAspThrProArg 956
    : ||| :||| ||| :||| ||| :||| |||
2373 AGCAGCATGTCA..... 2384
956 gLysLeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuVal 973
    : ||| :||| ||| :||| ||| :||| |||
2385 .....GACCAGACGCTCGGGTGA 2403
973 LysSerValGlyLeuGluValIleThrArgValHisLysAspValGlu 989
    : ||| :||| ||| :||| ||| :||| |||
2404 AGGGGGCAGCACTGAATACTTACCAACGATCGTCAACGATGTAAATG 2453
990 AlaGlnHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeu 1006
    : ||| :||| ||| :||| ||| :||| |||
2453 ..... 2453
1006 rLeuValAspArgGlyPheValPheSerLeuValArgAlaHisTyrLys 1023
    : ||| :||| ||| :||| ||| :||| |||
2454 .....GTGTT..... 2459
1023 lValAlaIleThrArgLeuGlnSerSerProAsnProAlaAlaLeuThr 1039
    : ||| :||| ||| :||| ||| :||| |||
2460 .....GATCCCAAGAGCTCAGCAAA 2480
1040 LeuArgMetGluPheThrArgIleLeuCysSerHisGlnHisTyrValTh 1056
    : ||| :||| ||| :||| ||| :||| |||
2481 ATGTTTACTGATTC.....AT 2497
1056 rLeuAsnLeuProCysCysProLeuSerProProAlaSerProSerPro 1073

```

2498 CCGAATGTTCC..... 2510
1073 eValSerSerThrThrSerGlnSerSerThrPheSerSerGlnAlaPro 1089
2510 2510
1090 AspProLysValThrSerMetPheGlnLeuSerGlyProPheArgGlnGln 1106
2510 2510
1106 nhISpHeuAlaGlyLeuLeuLeuThrGlnLeuAlaLeuAlaLeuGlnP 1123
2511ATGGGCTGCTGCGACATCCAGAACTCTACTGCTTG.... 2546
1123 roGlnAlaGlnGlnAlaPheLeuLeuHisLysLysAlaLLeSerAlaVal 1139
2547ATCGAAATCGTC 2558
1140 hISer...LeuLeuCysGlyHisAspThrAspProArgTyrAlaGlnAl 1155
2559 CACAGTGACCTCTTCACACAGCATGAC..... 2585
1155 arThrValLysAlaArgValAlaGlnLeuTyrLeuProLeuLeuSerLLea 1172
2586TGCAGAGAGATCGCTGCCATGATGAC..... 2614
1172 laArgAspThrLeuProArgLeuHisAspPheAlaGlnGlyProGlnGln 1188
2614 2614
1189 ArgSerArgLeuAlaSerMetLeuAspSerAspThrGlnGlyGlnLys 1205
2615CGATCAGCTCAAGTACCATCTGGAGACACAGAGACCTGGAGG 2658
1205 pLLeaGlyThrLLeaSnProSerValAlaMetAlaLLeaGlyGlyP 1222
2659 CCGCTGCTCAGCTGCTCAGCCACAT.....CTGGAGGTG 2693
1222 roLeuAlaProGlySerArgAlaSerLLeSerGlnGlyPro..... 1235
2694 CTGTACAGAGAGACGTGGGCGAACCCAGAGCAGCTCCAGATTATCAT 2743
1236Pro..ThrAlaSerArgAlaGlyCysAlaLeuSerAlaG 1248
2744 GGAGAAACTTCTCCGGACGCTGAACCGACGTCATTTCATGGGACGAG 2793
1248 luserSer.....ArgThrLeuLeuAlaCysValLeuTrpValLeuLys 1262
2794 ATTTCTGAACCATTTGGAAACTGCTGCTGCATGACAGCTATTATTACGA 2843
1263 AsnThrGlnProAlaLeuLeuGlnArgTrpAlaThrAspLeuThrLeuPr 1279
2844 CAAATGGAA..... 2852
1279 oGlnLeuGlyArgLeuLeuAspLeuLeuTyrLeuCysLeuAlaAlaPheG 1296
2853GATTACCATTTATGCCCACTGATCAAG.... 2879
1296 lutyTrLysGlyLysLysAlaPheGlnArgLLeaSnSerLeuThrPheLys 1312
2880ACTTTGGG 2888
1313 LysSerLeuAspMetLysAlaArgLeuGlnGlnAlaLLeuGlnGlyThrL 1329
2889 AAA.....ATGAGGACTGATGTGTAGATTCTCTAATGGAAACATT 2929
1329 eGlyAlaArgGlnGlnLeuValAlaArgSerArgGlnArgSerProPhe. 1345
2930 CATCATGTTTAAGAACCTCATT.....GGAAAGACGTTTACCCTTCG 2973
1346GlyAsnProGlnAsnValAlaGlyTrpArgLysSerVal 1357

2974 ACTGGGTGATCATGACATGCTGCAAAATAAAGTCTTCTCGAGCAATT 3023
1358 ThrIStrLysGlnThrSerAspArgValAspLysThrLysAspGlnLe 1374
3024 AATCAGTATGAGATATGCTGAACAA.....AAATTCTGTGATCAAGC 3067
1374 tGlnHisGlnAlaLeuValGlnGlyAsnLeuAlaThrGlnAlaSerLeuY 1391
3068 CAACCTTGAGCTACAGCTGTGAACAACACTTTCACCTGGCT.....G 3111
1391 aValLeuAspThrLeuGlnLLeLeValGlnThrValMetLeuSerGln 1407
3112 TTGCTTCTTATCTCAAGAGTCCCTGCACACTGGAAATTTTCAAGTGC 3161
1408 AlaArgLysSerValLeu.....G 1414
3162 AAGAGAGCCAAATCTTAACAAGTACGAGATATGAGAGACAGATTGG 3211
1414 yAlaValLeuLysValValLeuTyrSerLeuGlySerAlaGlnSerAla 1431
3212 CTTTGAATCAGACATGTGTCAACCTTGT..... 3245
1431 eupHeuGlnHisGlyLeuAlaThrGlnArgAlaLeuValSerLysPhe 1447
3246CAACACAGATAAAGTTTCATTCCAGAAATGTTGGCCCAATA 3287
1448 ProGlnLeuLeuPheGlnGlnLysPThrGlnLeuCysAlaAspLeuCysLe 1464
3288 TTAGAATGACATTATTCGCCGAGCGAGCTG..... 3320
1464 uArgLeuLeuArgHisCysGlySerArgLLeSerThrLLeArgThrHis 1481
3321CCGMAAGCCACCATC..... 3335
1481 laSerAlaSerLeuTyrLeuLeuMetArgGlnAsnPheGlnLLeGlyHis 1497
3336CCCATCTTCTTTATATGATGACAGTGTGAATTCATTCACCCGA 3380
1498 AsnPheAlaArgValLysMetGlnValThrMetSerLeuSerLeuVal 1514
3381 ACCTTCMAAATGTTGAATAATGATCATCACCAAGCTGATCATGAAGT 3430
1514 l.....GlyThrThrGln.....AsnPhe 1521
3431 CCAAGAGGCGAGAGAGAGACAGACATACAAAGTATTATTGATAAATCC 3480
1521 eGlnGlnLLeuArgSerLeuLysThrLLeLeuThrTyrAlaGln 1537
3481 TTCTGGAACACTGCCAGAG.....CAC 3503
1538 GluAspMetGlyLeuArgAspSerThrPheAlaGlnGlnAlaAspLe 1554
3504 AAATACCTCGCAAAACAGAGAAACTTTTGTAATAACTGTGTGCGCTT 3553
1554 lMetPheAsnLeuHisMetLLeLeuThrAspThrValLysMetLysGln 1571
3554 AATGAAAGGCTTTGGATTATGAACCATCATG.....C 3588
1571 lSglnGlnAspProGlnMetLeuLLeAspLeuMetLysArgLLeaArg 1587
3589 ACGACGAGAACAAAGAAACCGCATGAGCTGACCCGTCATGTGCTGAT 3638
1588 GlyTyrGln.....GlySerProAspLeuArgLeuThrTrpLeuGlnAs 1602
3639 TTCTACAAGAAATTTGAAGAGAGAAATGTTATTAAGGATTTGTACAA 3688
1602 nhMetAlaGlyLysHisAlaGlnLeuGlnLysnhHisAlaGlnAlaGlnInc 1619
3689 GCTGTGACCTGCGACAGAGAGTGTATTAAGTACACCGAAGGCGCTTACA 3738
1619 yMetValHisAlaAlaAlaValAlaGlnTyrLeuAlaLeuGlnGln 1635
3739 CATTGCTT.....CTCCATGCAAGGCTTCTTAAGTGTGCGAG 3776

```

1636 AspGlnArgHisLeuProValGlyCysValSer..... 1646
1637 GATGTG.....TGTGTGGCCCACTCAACCGCGGA 3808
1647 ....PheGlnAsnIleSerSerAsnValLeuGluIleSerAlaIleSerA 1662
1648 ..... 1662
1649 CGGTACACAGCCACCGACGAGCAGCAGTCAAGCAGCTTCACAG 3858
1662 sPaSPILeUSeRProAspGluGlyPheCysSerGlyLysHisPhe 1678
1663 ..... 1678
1664 AATTCATC.....CACTACTTGACAAGGCAAGATGTGG 3893
1679 ThrGluLeuGlyLeuValGly.....LeuLeuGluGlnAlaIleGlyTyr 1693
1680 GAGAGAGCCATTCCTTGGCAGAGCAGTACCGCAGCAG..... 3932
1693 rPheIleMetGlyGlyLeuTyrGlnAlaValAsnGluValTyrLysAsnL 1710
1694 .....TATGAG.....AACGAATGTTT..... 3950
1710 euIleProIleLeuGlnAlaHisArgSPyrLysLysLeuAlaVal 1726
1727 .....GATTATGACCACTCAAGCAGATTCG 3974
1727 HisGlyLysLeuGlnAlaPheThrLysIleMetHisGlnSerSerG 1743
1728 ..... 1743
1729 CTGMAAAACAGGCTCAGTTTATGMAAACATGTCACAAAGTATC..... 4019
1743 yTrpGluArgValPheGlyThrTyrPheArgValGlyPheTyrGlyAlaH 1760
1744 .....AGGCCCAAGCTGACTATTTCTGTGGTACTACGCAAG 4062
1760 IsPheGlyAsp...LeuAspGluGlnLupheValTyrLysLuproSer 1775
1761 ..... 1775
1762 GGTTCCTCCATCTCGCGGAAAGTTTCATTACCGAGGAAG 4112
1776 IleThrLysLeuAlaGluIleSerHisArgLeuGluLupheTyrThrG 1792
1777 ..... 1792
1792 TATGACCCCCGGAGATTTTGAAGCTCGCTTAACTCACTTCCAAA 4162
1793 ..... 4162
1794 uArgPheGlyAspSprValAlaGluIleLysAspSerTyrProValA 1809
1795 ..... 1809
1809 sPlySerLysLeuAspSerGlnLysAlaTyrLleGlnIleThrTyrVal 1825
1810 ..... 1825
1811 GCGAGATTTTAAAAAATCTCTGCGCATATATCAGCTTCACAGTG 4244
1826 GluProTyrPheAspLhr.....TyrGluLeuLysAspArg 1837
1827 ..... 1837
1828 AAGCCCAACTCGATCTGCTCTAGTTTCAACAGCCAGGTCTGAGACA 4294
1837 gValThrTyrPheAspArgAsnTyrGlyLeuArgThrPheLupheCysT 1854
1838 ..... 1854
1839 GATTTGAAGTTTTCAGGGTGAAGAGGTCCAGGATTTGAATTTCTC 4344
1854 hrProPhe.....ThrProAspGlyArgAlaHisGlyLys 1865
1855 ..... 1865
1856 GGCATATCCGAGAGGAGAGAAACCCAGAC.....AATGAA 4382
1866 LeuProGluGlnHisLysArgLysThrLeuLeuSerThrAsnHisAlaP 1882
1867 ..... 1882
1868 TTTCGAATATGTGATGAGAGAAACATATATCACTGATTAATTAAT 4432
1882 eProTyrIleLysThrArgIleArgValLysHisArgGluIleThrVal 1899
1883 ..... 1899
1884 ACCTGAATTTTAAAGTGTGAGTCAAGTCTGTTTCAATGTTGAAA 4442
1899 euThrProValGluValAlaIleGluAspMetGlnLysLysThrArgGlu 1915
1899 ..... 1915
1899 TCAGCCCTCGAGAAATGCCATTCGAGACATGACGAGCAAGCAAG 4532

```

```

1916 LeuAlaPheAlaThrGluGlnAspProAspAlaLysMet..... 1929
1917 ..... 1929
1918 ATCAACACATGCTGACACACACCTGGATGATGACCCAGCTGCCATCA 4582
1930 ....LeuGlnMetValLeuGlnGlySerValGlyProThrValAsnGln 1945
1931 ..... 1945
1932 CCCGCTTCATGCTCTCGAAGCGATGCTGACCGACCTGTCATGGGG 4632
1945 yProLeuGluValAlaGlnValPheLeuAlaGlu.....IleProGlu 1959
1946 ..... 1959
1947 GCTTGCAAACTACGMAAGCCCTTTCACAGACCGGACTGTCAGAG 4682
1960 AspProLysLeuPheArgHisAsnLysLeuArgLeuCysPheLysAs 1976
1961 ..... 1976
1962 CACCTGAGGCCCATGAAGAATGAGAACTC.....AAGA 4720
1976 pPheCys..... 1978
1979 ..... 1978
1979 CCGTTCACGAGATGATGATGATGATGATGATGATGATGATGATG 4720
1979 ..LysLysCysGluAspAlaLeuArgLysAsnLysAlaLeuIleGlyPro 1994
1995 ..... 1994
1995 GAGCAAAAGTCAGGAGGACTGAGG..... 4796
1995 AspGlnLysGluTyrHisArgGluLeuGluArgAsnTyrCysArgLeuAr 2011
1996 ..... 2011
1997 .....CGTTCACGAGATGATGATGATGATGATGATGATGATGAA 4837
2011 gGluAlaLeuGln 2015
2012 ..... 2015
2012 GMAAAGGTGAG 4850
seq_name: /cgn2_6/prodata/2/lna/6A_COMB.seq:us-08-923-137-2
seq_documentation_block:
; Sequence 2, Application US/08923137
; Patent No. 6083716
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Farina, Steven F.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Chimpanzee Adenovirus Vectors
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr., P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,137
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,700
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GAYVN.021CIP1USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 36519 base pairs
; TYPE: nucleic acid

```

```

; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-923-137-2

```

```

Alignment_scores:
  Quality: 191.00      Length: 1635
  Ratio: 0.302        Gaps: 88
  Percent Similarity: 38.716      Percent Identity: 21.101

```

Alignment_block:

US-09-736-968a-2 x US-08-923-137-2/reverse ..

Align seg 1/1 to reverse of: US-08-923-137-2 from: 1 to: 36519

```

125 ArgArgTyrGlnTyrLeuSerAlaIatYserProValThrThrAspTh 141
    ||||| : : : : : ||| : : : ||| : : :
8593 GGTGGCTACGCGCAGTCTTCTCCGGGCGACGTACCAAGAACCTACCGCGC 8544
    ||| : : : : : ||| : : : ||| : : :
141 rGlnArgGlnArgGlnLysGlyLeuProArgGlnValPheGlnLysAsp 158
    : : : : : ||| : : : ||| : : :
8543 GCGGGCGGCGAGTGCCTCCGCGCCCT..... 8516
158 lAserGlyAspGlnArgSerGlyProGlnAspSerAsnAspSerArgArg 174
    ||||| : : : : : ||| : : : ||| : : :
8515 ..... GCTCCGGGCGCGAGCGCCCTGCTCCGACCGCGCGCGC 8480
175 GlySerGlySerPro.....GlnAsp..... 181
    : : : : : ||| : : : ||| : : :
8479 GGTCTCTGCGCGCTGCTTCAACCATGGAAGAGCGACCGCCCGCTGCC 8430
182 .....ThrProArgSerSerg 187
    ||||| : : : : : ||| : : : ||| : : :
8429 CCAGCGCGCCCAAGAGAAGACGGGGAGCGGTGTACACCCAGGGGAGCAG 8380
187 lYAlaSerSerTlPheAspLeuArgAsnLeuAlaAlaAspSerLeu... 202
    ||| : : : : : ||| : : : ||| : : :
8379 GGAACCTCAAGCCATCGACGTGCGCACCAATGGCGGTGAGATCAAG 8330
203 .....LeuProSerLeuLeuGln.....Ar 209
    ||||| : : : : : ||| : : : ||| : : :
8329 TACCAATCTGGACCTCCGCGCGCCCTGGAAAAAATCTCCGCAAGTCAACCG 8280
209 gAlaAlaProGlnAspValAsp.....ArgArgAsnGlnThrLeuA 223
    ||| ||| : : : : : ||| : : : ||| : : :
8279 GCGCGCGCGCGCTGCCACTGACTGCTCCGAGGCTCTCAGAGACCTGG 8230
223 rGArgGlnHisArgProProAlaLeuLeuThrLeuTyrProAla..... 237
    : : : : : ||| : : : ||| : : : ||| : : :
8229 ACAGCTCGGCGCTGCGCGCCCTGCTCTGCTCCGACCGCGCTCGCGC 8180
237 ..... 237
8179 GAGGTCTGACTCTCCCTCCGCGCGGCGCTGTCAGCATGACCACCATCGA 8130
238 .....ProAspGlnAspGlnAlaValGlnArgCysSera 249
    ||| : : : : : ||| : : : ||| : : :
8129 GGCAGAGAGGCGCAGGCGGACACACAGACGTAGTACAGCAC.....G 8086
249 rGProGlnProArgGlnHisPheGlyGlnArgIleLeuValLysCys 265
    : : : : : ||||| : : : ||||| : : :
8085 AGATGCAAGCGCGCACGACTCCACTTCCACTGAATTCCTCGTCAAGA 8036
266 LeuSerLeuLysPheGlnIleGlnIleGlnProIle..... 277
    : : : : : ||| : : : ||| : : : ||| : : :
8035 ACTCAGGTACACTCTGTCAGACAGCATGCCGTCCGTCAGCGTGGAGCA 7986
278 .PheGlyIleLeuAlaLeuTyrAspValArgGlnLysLysIleSer. 293
    ||| ||| : : : : : ||| : : : ||| : : :
7985 CTGTGCGCGCTTG.....TATTAACACAAGCAGCAGTGTCTCG 7948
294 .....GlnAsnPheTyrPhe...AspLeuAsnSeraSperMetLys 306

```

```

7947 CGCGCGCTGCGGATTTCTTACTTCATCATCATCAACAGCATTCCTCAT 7898
    : : : : : ||||| : : : ||| : : :
307 GlyLeuLeuArgAlaHis.....GlyThrHisProAlaIleSe 319
    : : : : : ||| : : : ||| : : :
7897 TGGTGGCAGAGATCCAGTTCTTCCCATCGGCTGCATCTCGACCGCA 7848
319 rThrLeuAlaArgSeraAlaIlePheSeraValThrTyrProSerProAspI 336
    : : : : : ||| : : : ||| : : :
7847 AAGCTC..... 7841
336 lPheLeuValIleLysLeuGlnLysValLeuGlnGlnLysPleSer 352
    ||||| : : : : : ||| : : : ||| : : :
7840 ..TTCCTACCTACGATGTGGAAACCTACACCTGATGGGTCCTTCGCG 7793
353 GLucyScyGlnProTyrMetValLeuLysGlnValAspThrAlaLysAs 369
    : : : : : ||||| : : : ||||| : : :
7792 AACACACTGCTCCCTTCATCTGCTGATCAAGAAC.....TCGG 7752
369 nLysGlnLysLeuGlnLysLeuArgLeuAlaIleGlnIlePheCysThrA 386
    : : : : : ||| : : : ||| : : :
7751 GACAGACCGGCTGCTGACCTCGCCCTC..... 7724
386 rGLeuGlyArgTyrArgMetProPheAlaThrPheAlaValHisLeuAla 402
    : : : : : ||| : : : ||| : : :
7723 .....GATCTGCCCTCCAGCTCAAAAGGATGCTGCGCAC..... 7688
403 AsnIleValSerSeraIleGlnLeuAspArgAspSeraSperSergL 419
7688 ..... 7688
419 yGlnArgArgProAlaThrThrAspArgArgArgGlyProGlnAspA 436
7688 ..... 7688
436 rGlnAserSerGlyAspAspAlaCysSerPheSerGlyPheArgProAla 452
    ||| : : : : : ||| : : : ||| : : :
7687 .....GGGAGACCCCGCACCTTCTACGCGCGACACCCCGAA 7652
453 ThrLeuThrValThrAsnPhePheLysGlnGlnAlaGlnArgLeuSerAs 469
    : : : : : ||| : : : ||| : : :
7651 AAATGCGCGTGGCGAGCAGTTCCGTCAGTACCGGAGCGCTCGCAAC 7602
469 pGln.....AspLeuPheLysPheLeuAlaAspMetArgArgProS 483
    ||||| : : : : : ||| : : : ||| : : :
7601 CGCCCTGGCGCGTGGATCTGG.....A 7579
483 eSerLeuLeuArgArgGlnArgProValThrAlaGlnLeuLysIleAsp 499
    : : : : : ||||| : : : ||||| : : :
7578 CTTCCTTCTCCCG..... 7565
500 lLeSerProAlaProGlnAsnProHisPheCys..... 510
7564 .....GCCAACCCCATCTGCGCGACTGGGCGCTCGAGCA 7530
511 .....LeuSerProGlnLeuLeu.....HisIleLysProI 521
    : : : : : ||| : : : ||| : : :
7529 GCACGGGCTCAGCGACCGCGAGCTCACCTACGAGAGCTCAAGAAC 7480
521 yRProAspProArgGlyArgProThrLysGlnIleLeuGlnIlePheProAla 537
    ||| : : : : : ||| : : : ||| : : :
7479 TGCCCAACGTCAAGGCGCTCG.....CGATTCTGGAA..... 7445
538 ArgGlnValTyrAlaProHisThrSerTyrArgAsnLeuLeuTyrValTy 554
    : : : : : ||||| : : : ||||| : : :
7444 .....CTCTACATCGT 7434
554 rProHisSerLeuAsn.PheSerSeraArgGlnLysSer..... 566
7433 GGGCACAACATCAACGCTTCCAGCAGATCGTCTGCGCCAGGTCA 7384
567 .....ValArgAsnLeuAlaValArgVal.GlnTyrM 577
    ||||| : : : : : ||| : : : ||| : : :

```

7383 TCACACACCGGCGAGGTGCCCAACCTTTCCGCATCATCGCAACTTT 7334
577 eethr.....GlyluAspProserGln..... 584
7333 ATGCCCGCGGGGCAAGATCCTCTTCAACAGATGTTACTTTGGCTTGGC 7284
585AlaLeuProValIlePheGlyLysSe 593
7283 CAACCCCGCTACAGAAAGCCAGGACCTTCAGCTCG...GGAGCAGG 7237
593 rSerCysSerGluPheThrArgGluAlaPheThrProValValThrHisA 610
7236 GGGCTGGGAGGATATCGACTTCAGCA.....CCAGTCTCTAAGGTC 7193
610 snLysSerProGluPheTyGluGluPheLysLeuHisLeuProAlaCys 626
7192 ATGTCAGGAGACATTTCGCCCTCAC.....GCACACCTCCCGCG... 7152
627 ValThrGluAsnHisLeuLeuPheThrPheThrHisValSerCysGI 643
7151CA 7150
643 nProArgProGlyThrAlaLeuGluThrProValGlyPheThrTrileP 660
7149 AGGCGCGCCAGGC.....CTACGCCCTCCCGCGAGAAAGGCTGCTGC 7106
660 roleu.....LeuGlnHisGlyArgLeuArgThrGlyPro 671
7105 GCCTACAGGCGCTCAACCAAGTCTCATGCTAGAGCTC..... 7068
672 PheCysLeuProValSerValAspGlnProProProSerTySerValLe 688
7067TTACCGTCAGATCAGAGCGGGTTCCCTCGAAGA..... 7032
688 uThrProAspValAlaLeuProGlyMetArgTrpValAspGlyHisLysG 705
7031GTACTGGAGGACCGCGAAGATTCTCTTCACACCGC 6995
705 LyValPheSerValGluLeuThrAlaValSerValHisProGlnAsp 721
6994 GAGCTGTGGAACAAAGGAGCACTCAAGTATGACATCCACGAGAC 6945
722 ProTyLeuAspLysPhePheThrLeuValHisValLeuGluGluGlyAl 738
6944 CCTGCACTA.....CTGCGCTCTGGACG 6922
738 aPheProPheArgLeuLysAspThrValLeuSerGluLysValGluG 755
6921 TCCTAGTCACCGC..... 6909
755 InGluLeuArgAlaSerLeuAlaAlaLeuArgLeu..... 766
6908CGAGCTCGTCCCAAGCTGCAGACTCTACGCCCACTTCAT 6867
767AlaSerProGluProLeu.....ValAlaPheSe 776
6866 CCGGAGACTCGGTGGAGCTGCCACGCGCCACTTCAACATCTTCCAAGGC 6817
776 rHisHisValLeu.....AspLysLeuValA 785
6816 CCACCACTCTCTCCAACTGCGCATGTTTCCGCCAGATCGCTA.... 6771
785 rGluLeuValIleArgProProIleIleSerGlyGlnIleValAsnLeuGly 801
6770CCGCGCGGAAAGCCAGTCGACCAACCTGGG 6738
802 ArgGlyAlaPheGluAlaMetAlaHisValAlaSerLeuValHisArgSe 818
6737 CCGGGGGCTCTGCGCCCTCGCAGAGCTGT.....TG 6703
818 rLeuGluAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAlaT 835
6702 ACTACGTGGCGCGCCAGCATCGCGGGG.....GCGCTGTACATCCC 6662
835 yValHis..TyrAlaPheArgLeuProGlyThrGlu..... 846
6661 ACCTACATCGGCATCTCGAGGAGCCCTCTATACAGATCTGCGG 6612
847ProSerLeuProAspGlyAlaProProValThrValGln 859
6611 CATGTACGGCTCGCGCTTACCATCCCATGGCTTGGGCGACGCCCTCA 6562
860 AlaAlaThrLeuAlaArg.GlySerGlyArgProAlaSer..LeuTyTrLeu 875
6561 GTCCCTACAGAGCGCGCTGCGCGGAAATGGCAACCTGCCCGGAC 6512
876 AlaArgSerLysSerLysSerSerAsnProAspLeuAlaValAlaArg 892
6511 GATCTGGCCACTTCCATCGACTACTTTCACCCCGACCTCTCCCGGCAT 6462
892 ogLysSerValAsp...AspGluValSerArgIleLeuAlaSerLysLeuL 908
6461 CTTCACCATCGATGCGCGACCCCGCGAGGTCATGCTGACCCCTG. 6413
908 euHisGluGluLeuAlaLeuGlnTrpValValSerSerSerAlaValArg 924
6413 6413
925 GluAlaIleLeuGlnHisAlaTrpPhePhePheGlnLeuMetValLysSe 941
6413 6413
941 rMetAlaLeuHisLeuLeuLeuGlyGlnArgLeuAspThrProArgLysL 958
6412CCCCCTTCTGCT 6400
958 euArgPheProGlyArgPheLeuAspArgIleThrAlaLeuValLysSer 974
6399 CCGGCAAGGGCGGAGCGCTCTGCTGACTATAGACCCCTGCGCGGAG 6350
975 ValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaGI 991
6349 GTGGCCACCACTGTG..... 6335
991 uHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeuY 1008
6334GACCTCATACCCCTGC 6319
1008 aLAspArgGlyPheValPheSerLeuValArgAlaHisTyLysGluVal 1024
6318 ATAATCGGGGCTGCGAGGTCAGAAATCGTC..... 6290
1025 AlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeuAr 1041
6289CCGACGAGCTGACACCGTC.. 6269
1041 gMetGluPheThrArgIleLeuGlySerHisGlnHisTyTrValThrLeuA 1058
6268TTCGCCGAATGAATGCTGCGCGGAGTACGTGACGTCA 6226
1058 snLeu.Pro.....CysCy 1062
6225 ACATGCGCGCCAGAGGCGCGCCAGCAAGAAAGAAACCAACCATGCGC 6176
1062 sProLeuSerProProAlaSerProSerProSerVal..... 1074
6175 TCCA...TCGCCAAGCTCTCTCAACGCGCTTACGAGCTCTTGGCCAC 6129
1075 ..SerSerThrThrSerGlnSerSer.....ThrPheSer 1085
6128 CAAGCTCGACAAACAAAGATGCTTTTCTGACCAATGACGAGGGGC 6079
1086 SerGlnAlaProAspProLysValThrSerMetPheGluLeuSerGlyPr 1102
6078 TCTTGAAAGGATCTCGCGCGGACCGTCA.....ATATCAATTC 6038

```

1102 ophAaTgGlnGlnHisPheLeuAlaGlyLeuThrGluLeuAla 1119
      |||
      |||
6037 TCCTCGTTT.....CTAGAACTGACAACTGAGTG 6007
1119 euAlaLeuGluProGluAlaGluGlyAlaPheLeuHisLysLysAla 1135
      |||
      |||
6006 CCGAGGTCATGCCGCTTCGAGAGGGAATCCTACCCCAACAGCTGGCG 5957
1136 lAsSerAlaValHisSerLeuLeuCysGlyHisAspThrAspProArgTy 1152
      |||
      |||
5956 CTCCTG.....GACAGCATCCGGAAGA 5934
1152 rAlaGlu.....AlaThy 1157
      |||
      |||
5933 CAGTAGAGAGAGAGAGGACCCGACCTTTTATACCCCGCGCGGA 5884
1157 aLysAlaArgValAlaGluLeuTyLeuProLeuSerIleAlaArg 1173
      |||
      |||
5883 CCCCCGCTACGTGGCTTACACTTACAGCCAAATCACCTTCCTGGATGTG 5834
1174 Asp.....ThrLeuProArgLeuHisAspPheAl 1183
      |||
      |||
5833 GACAGAGGGGACATGTGTTCATACCTCGAAAGGTGAGACCGGTGT 5784
1183 aGluGlyProGlyGlnArgSerArgLeuAlaSerMetLeu..... 1196
      |||
      |||
5783 GGACAAAGACCGCTACCCGTCCACGTGGCTCTTGTGTGGCTGGA 5734
1197 .....AspSer 1198
      |||
      |||
5733 CCGCGGCTTCGTCTCGAGTGGCGGGGTTCCTCTACGAGAGACCGC 5684
1199 AspThrGluGlyGlu.....GlyAspIle.A 1207
      |||
      |||
5683 GGCACACCGCTCGAGACCGGCCATTAAGTGGGTACGGGACACGGA 5634
1207 laGlyThrIleAsnProSer..... 1213
      |||
      |||
5633 CAGCTCTTTGTCTACCAAGGGGAGACAGCTCATGAGACCAAGTA 5584
1214 .....ValAlaMetAlaIleAlaGly.....ProLeuAl 1224
      |||
      |||
5583 AGAAAGCATCAAAAGACGGCGGAACGTGTTTGAACCGGACAG 5534
1224 aProGlySerArgLaserIleSerGlnGlyPro.....ProThrLas 1239
      |||
      |||
5533 CCGGACCTCACCTGTCTGTGAGTGGAGACCTCTGCGCCACTCGG 5484
1239 eATGAlaGlyCysAlaLeuSerAlaGluSerArgThrLeuLeuAla 1255
      |||
      |||
5483 CGCGGAC.....GCCTAGCCCCCGAGTCC.....GTCTCTCTGTC 5447
1256 CysValLeuThrValLeuLysAsnThrGlnProAlaLeuGlnArgTy 1272
      |||
      |||
5446 CCGAAGCTCTACGCCCTCAAGTCC..... 5433
1272 pAlaThrAspLeuThrLeuProGlnLeuGlyArgLeu..... 1284
      |||
      |||
5422 .....CTCCTCTGTCTCCGCTGGCGGACAGACTTCCAAAGTAAGC 5383
1285 .....LeuAspLeuLeuTy 1289
      |||
      |||
5382 TCCGCGCCAAAGGCCACGCGGCCGACGCTCAACTGCAACTATGATGTC 5333
1290 LeuCysLeuAlaAlaPheGluTyLysGlyLysLysAlaPheGluArg 1306
      |||
      |||
5332 AATGCTACTCGGCC...GACGGCAGAGGCGCATGG...GAGCGGT 5289
1306 eAsnSerLeuThrPheLysSerLeuAspMetLysAlaArgLeuGlu 1323
      |||
      |||
5288 TTGCGAC.....AGCCGATGAGCTGAAGCGACACCTCCCA 5251
1323 lValAlaIleLeuGly..... 1327

```

```

5250 GCGCCAGCCCGCGGACCCCTTCACCTGTGAGGAGACCACTGACG 5201
      |||
      |||
1328 .....ThrIleGlyAla.....ArgI 1333
5200 CGGAGCTCCGACCTCGAAGAGACCGGACCGCTGGCGCGGAGCCCA 5151
1333 nGluMetValArgArgSerArgGluArgSerProPheGlnAsnProGluA 1350
      |||
      |||
5150 TCGTGTGGCCCTACTCCGCGACCGTCC.....AACCCGCA 5110
1350 sn.....ValArgTyr..ArgLysSerValThrHisTrpLysGlnThr 1364
      |||
      |||
5109 ACGAGAGGTCTGTGATGAGATGCCGTAGAGCACAATCACCAAGCTCT 5060
1364 eAspArgValAspLysThrLysAspGluMetGlnHisGluAlaLeuVal 1380
      |||
      |||
5059 GGGACCGCTGAGACTCTTGCAACAAACCTCTCCAAATGGCCATGGCC 5010
1381 GluGly.....AsnLeuAlaThrGluAlaSerLeuVal 1392
      |||
      |||
5009 GACGCACTCAACCCGCTGAAAACTTGGCTGCTCCAGAGACTCTATC 4960
1392 lLeuAspThrLeuGluIleIleValGlnThrVal..... 1403
4959 CCTGGGGGAGAGCCGCTCTGCGGAACTGTGCCGAGAAATCCAG 4910
1404 .....MetLeuSerGluAla.....ArgGluSer..... 1411
4909 TGGCGAGATGATGAACAGGTGGCCGCCCTCTCCGGAGAGACGACG 4860
1412 .....ValLeuGlyAlaValLeuLysValValLeuTy 1422
      |||
      |||
4859 TGTCTCTCCCTCAACTACCACTGCAGCCGCTATCGGGGTATCTACGG 4810
1422 rSerLeuGlySerAlaGlnSerAlaLeuPheLeuGlnHisGlyLeuAlaT 1439
      |||
      |||
4809 CCGCAGCGGCTGGGCAAGTCCAGCTG...CTCCGAACTGTCTTCGG 4763
1439 hGlnArgAlaLeuValSerLysPheProGluLeuLeuPhe 1452
      |||
      |||
4762 CCCAG.....CTCATCTCCCCCGCCCGAAACCGTGTTC 4728

seq_name: /cgn2_6/plodata/2/lna/5B_COMB.seq:us-08-804-227C-7
seq_documentation_block:
; Sequence 7, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostek, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.

```



```
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830
US-09-804-227C-7
```

```
alignment_scores:
Quality: 182.00      Length: 1649
Ratio: 0.285        Gaps: 89
Percent Similarity: 38.690      Percent Identity: 21.953
```

alignment_block:

US-09-736-968a-2 x US-08-804-227C-7 ..

Align seg 1/1 to: US-08-804-227C-7 from: 1 to: 44377

```
2  ALaAlaSerGluArgAlaPheAla.HisLysIleAsnArgThrValA 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23189 GCGCTGCTCGACGAGATCAACGCGACCGCGCTCACCGCGCTC 23238
18  LaAlaGluValArgLysGlnValSerArgGluArgSerGlySerProHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23239 TCCTCT.....GCTCGACTGTGCACCGACGCGCGCGACCCCGC 23282
35  SerSerArgCysSerSerSerLeuGlyValProLeuThrGluValA 51
:::|||||:|||||:|||||:|||||:|||||:|||||:
23283 CGTGGCGCTGCTC..... 23294
51  LgluProLeuAspPheGluAspValLeuLeuSerArgProProAspAlaG 68
|||||:|||||:|||||:|||||:|||||:|||||:
23295 .....CGTCACCGCGCACCCCTCGCGCTC 23316
68  LuProGluProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
:::|||||:|||||:|||||:|||||:|||||:|||||:
23317 GTCCAGCGCTTGGCGCA..... 23333
85  LeuLeuLeuGluProArgGluCysArgThrThrGluProGlyTyleProly 101
|||||:|||||:|||||:|||||:|||||:|||||:
23334 .....CTTCGGCAGCAGCCGCGACCGCT..... 23354
101  sAspGluLysLeuAspAlaGlnValArgAlaValAlaGluMetTyrTleG 118
:::|||||:|||||:|||||:|||||:|||||:|||||:
23355 .....GTGACCGCTGACCTCGCGCGCT.....CGCACCGCGC 23388
118  LuAspTyrValIleValHisArgArg..... 126
:::|||||:|||||:|||||:|||||:|||||:|||||:
23389 CCGGAGAACTGCGTGCACCGCGGTGCCACGCTGTGGGCGCTGGCGCG 23438
127 .....TyrGlnTyrTle 130
:::|||||:|||||:|||||:|||||:|||||:|||||:
```

```
23439 GGTGGCGCGCTGTGACCTGCCGAGGTGTGGGCGCGCTCATCGACTTC 23488
130  uSerAlaAlaIleTyrSerProValThrThrAspThrGlnArgGluGlnL 147
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23489 CCGCGCGCGCGCGACCGCGGTCTGTGACCGCTGTGCGCGCGCT..... 23531
147  ysgLysLeuProArgGluValPheGluGlnAspAlaSerGlyLyspsGluArg 163
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23532 .....CCTGCGCGCAACC.....CGCGCG 23549
164  SerGlyProGluAspSerAsnAspSerArgArg.....Glyse 176
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23550 CGAGGACAGATCGCGCTGACGATGCGGCGCGCTTTCGCGCGCGGTCC 23599
176  rGlySerProGluAspThrProArgSerSerGlyAlaSerSerIlePheA 193
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23600 TGGGGAACCGCGCGCGCTCCCGCGCGCG..... 23627
193  sPLeuArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGluArg 209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23628 .....GGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCT.....CATCGC 23660
210  AlaAlaProGluAsp.....ValAspA 217
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23661 CGGCGACCTACAGACGTGCGCGCGCGCGCGCGCTGCTCCCTCGAG 23710
215 ..... 217
23711 ACGCGCGCGCGCGCTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 23760
217  rGArgAsnGluThrLeuArgArgGlnHisArgProProAlaLeuLeuThr 233
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23761 GCGCGCGCGCGCGCTGACCGCGGTCTCCCTGCTCCCGCGCGCGCGCGAGCT 23810
234  LeuTyrPro.....AlaProAspGluAspGluAlaValGlu 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23811 CACGACCGCGCGCGCGCTGCGCGCGCGCTGCTGACGACGACGCGCGCACCG 23860
245  uArgCysSerArgProGluPro..... 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23861 TCGCGGTGACAGCGCGCGCGCGCTGTGCGCGCGCGCGCGCGCGCGAGCG 23910
253  ..ProArgGluHisPheGlyGluArg.IleLeuValLysCysLeuSerLe 268
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23911 GCACCGCGCGCGCAT.....CGCGCGCGCGCTCGCGCGCGCGCGCGCG 23957
268  uLysPheGluIleGluIleGluProIlePheGlyTyleLeuAlaLeuTyrA 285
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23958 CGGCGACCTGTGTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG... 24003
285  sPValArgGluLysLysIleSerGluAsnPheTyrPheAspLeuAsn 301
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24004 .....GTGCTGTCTCTCTGCTGCTGCTGCGAGTGTGG 24033
302  SerAspSerMetLysGlyLeuLeuArgAlaHisGlyThrHis..ProAl 317
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24034 GCGCGCGCGCGCGCGCGCGCGCTACGCGCGCGCGCGCGCGCGCGCGCGCG 24083
317  aIleSerThrLeuAlaArgSerAla.....IlePheSerValThr 331
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24084 GCTGGCGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCTCGTGGCT 24133
331  yRPro..... 332
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24134 GGAGCGCGCTGGCGCGGAGACGCGCGCGCGCGAGGTGCGGAGGCGGATTC 24183
333 .....SerProAspIlePheLeuValI 340
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24184 CTCAGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 24230
340  eLysLeuGluLysValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
24231 GACCTTGGCGCGCGCGCTGAGCGCGCG.....AGCGCTGCGGTGCGG 24274
```

```

356 .....GluProTyrMetValLeuLysGluValAspThr 367
      |||||
24275 TCGGCAGCGTCGAGTCGGAGCCGGTTCGCCCTCTACACCTGGGTGGCT 24324
      |||||
367 LAlYsAsnLYSGluLYSLeuGLULysLeuArgLeuAlaIaGLuInPhe 363
      : : : :
24325 CCCCCCTACTCTTCGACGACATCCCGGACGTGACGGCTGGCGGCG 24374
      : : : :
384 CysThrArgLeuGLYArgTyrArgMetProPheLatrPThrAlaValH1 400
      |||||
24375 CGAACTGCCCCCGACGACCGGACACTGAC.....CA 24406
      : : : :
400 sLeuAlaAsnIleValSerSerIaGLYInLeuAspArgAspSerAsps 417
      : : : :
24407 CCTC.....CGAACTGTCGGCGAGCTGACGGCGGCACTCCGGC 24444
      : : : :
417 eRGLUGLYGUAArgArgProAlaTrpThrAspArgArgGLYPro 433
      : : : :
24445 CACAAGGGGCGACGCCCTGCTGGTGGCGGCGCACACGCCGCGC 24494
      : : : :
434 GlnAspArgAlaSerSerGLYAspAlaCysSerPheSerGLYPheA 450
      : : : :
24495 CGTCTCGGACAGCTCTCGGGGACGGGTGAGCAGCGCCGCGCTTCC 24544
      : : : :
450 rGPProAla.....ThrLeuThrValThrAsnPheLYSGlu 463
      |||||
24545 GCGACCTGGGCTTGCACCTCGCTGACGCCCTCGAACTG..... 24582
      : : : :
464 AlaGLUArgLeuSerAspGLUAspLeuPheLYS.....PheLe 476
      : : : :
24583 CCGCAGCGGGCTGAGACGACGCGGGCTCAAACTGCCACCTCCCTGGT 24632
      : : : :
476 uAlaAspMetArgArgProSerSerLeuLeuArgArgLeu..... 489
      : : : :
24633 CTTCGACCACTCCAGCCGGCGGGCTGCGCCGACCTCGGTGAGGAC 24682
      : : : :
490 .....ArgProValThrAlaGLuInLeuLY 497
      |||||
24683 TCCTGGCGGGAAAGACACGCCGCGAGCGGGCGGCCGACACCCGGTA 24732
      : : : :
497 sIleAspIleSerProAlaProGLUAsn.....ProH 508
      |||||
24733 CGGACGAGGAGCCCATCGCATTCGGCATGGCTGCCGTCGCCGCCGG 24782
      : : : :
508 IsPheCysLeuSerProGLULeu.....HisIleLYSPRO 520
      |||||
24783 CGGGGTGACATGCCCGAGACCTGTGGACCTGTAACCGGTGGAGCG 24832
      : : : :
521 TyrProAspProArgGLYArgProThrLYSGluIleLeuGLUPheProAl 537
      : : : :
24833 ACGCATATACCCCTTCCGACCAACCGGGATGGACACAGACCCCTTC 24882
      : : : :
537 aArgGLU.....ValTyrAlaProHisThrSerTyrArgAsn 549
      : : : :
24883 TACGACCCCGACCCGACTGCCCGGAGCACCACTACGTGGCGGAGGG 24932
      : : : :
550 LeuLeuTyrValTyrProHisSerLeuAsnPheSerSerArgGLNLYSe 566
      : : : :
24933 .....CGGGTCTCTGCACGACGCGGGCG 24955
      : : : :
566 rValArgAsnLeuAlaValArgValGLN.....TyrMetThrGLYGLU 580
      : : : :
24956 AGTTCGACCCCGCTTTCGGCATACAGCCCCCGGAGGCCCTGCGCATG 25005
      : : : :
581 AspProSerGlnAlaLeuProValIlePheGLYLYSserSerCysSerGI 597
      : : : :
25006 GACCCGACACAGCGGCTGATCTCTG.....GAGAGCTCTGGGAGTTC 25046
      : : : :
597 uPheThrArgGLUAlaPheThrProValValTyrHisAsnLYSserProG 614
      : : : :
25047 CTTCGAACGGCGCGCATGACCCGCTCAACTGCGCGGACAC..... 25089

```

```

614 LuPheTyrGLUInLuPheLYSLeuHisLeuProAlaCysValThrGLUAsn 630
      : : : :
25089 ..... 25089
631 HisHisLeuLeuPheThrPheTyrHisValSerCysGLNProArgProGI 647
      : : : :
25090 .....CGCACCGG 25097
      : : : :
647 yThrAlaLeuGLUInThrProValGLYPheThrTrpIleProLeuLeuGLN 663
      : : : :
25098 GGTCTCTGGCGACCAAC...GGACAGACATACGTGCGGCTCTCCAGG 25144
      : : : :
664 .....HisGLYArgLeuArgThrGLYProPheCysLeu 674
      : : : :
25145 ACGGCGACGAGAACTTCGACGGCTACATCGCACCGGC..... 25182
      : : : :
675 ProValSerValAspGLNProProProSerTyrSerValLeuThrProAs 691
      : : : :
25183 .....AACTCGGACGAGTATGTCGGCGG 25208
      : : : :
691 rValAlaLeuProGLYMetArgTrpValAspGLYHisLYSGLYAlaPheS 708
      : : : :
25209 GCTCTCTAC.....GTCCTTG 25225
      : : : :
708 eValGLULeuThrAlaValSerSerValHisProGLNAspProTyrLeu 724
      : : : :
25226 GACTGGAGGGCCCCCGCGTAC..... 25248
      : : : :
725 AspLYSPhePheThrLeuValHisValLeuGLUGLYAlaPheProH 741
      : : : :
25248 ..... 25248
741 eArgLeuLYSAspThrValLeuSerGLUInValGLUGInGLULeuA 758
      : : : :
25249 .....GTGCACACCGCTGCT 25264
      : : : :
758 rGLAserLeuAlaIaLeuArgLeuAla..... 767
      : : : :
25265 CGGCTCTCTGGCGGACATGCACTGGCGGTGACATGCGCGCGCGC 25314
      : : : :
768 .....SerProGI 770
      : : : :
25315 GAATGCGACTACGCCCTCGCGGGGGCCACAGGTATGTCACACCCCGA 25364
      : : : :
770 uProLeuValAlaPheSerHisHisValLeuAspLYSLeuValArgLeu 787
      : : : :
25365 GATGCTGTGTGAGTTCGCCCT.....CAAGCGAG 25393
      : : : :
787 alIleArgProProIleIleSerGLYInIleValAsnLeuGLYArgGLY 803
      : : : :
25394 CGGTGTCGCCGAGAGCGCGACAGGCGTTGCGGAGCGGCGCGAGCGG 25443
      : : : :
804 AlaPheGLUAlaMetAlaHisValAlaSerLeuValHisArgSerLeuGI 820
      : : : :
25444 GTGCGTTCGCCGAGGAGCGGGATGCTGTGTGAGCGGTGCGGA 25493
      : : : :
820 uAlaIaGLAspAlaArgGLYHisCysProGLNLeuAlaIaTyrValH 837
      : : : :
25494 GCGCGACAGAG...AAGGCCAT...CGGTATGCGCGGTGTGCGGG 25534
      : : : :
837 IsTyrAlaPheArgLeuProGLY.....ThrGIUProSer 848
      : : : :
25535 GCACTGCGCTTACACAGAGAGGTGCGACAGACGGCTCACCGACCCAGC 25584
      : : : :
849 LeuPro.....AspGLYAl 853
      : : : :
25585 GGGCGCGCCAGCAGCGGATACGGAGCGCGTGGCGGACCGGAGGCT 25634
      : : : :
853 aProProValThrValGLNAlaIaThrLeuAlaArgGLYSerGLYArgP 870
      : : : :
25635 GACCGCGCGAGCTGTGAGCGGTGCGAG...GCGCAGGAGCACCGGACGC 25681
      : : : :
870 ro..... 870

```

```
25662 CGCTCGGCGACCCCATCGAGCGCGCGCTGTCGCCACCTACGCGCG 25731
      ||
871 .....AlaSerLeuTyrLeuAlaArgSerIleSerSe 883
      |||:::||||:::
25732 GACCGCGCGACGCGCGCTGCTGGTTCGCTGAG..... 25770
      |||:::
883 rSerAsnProAspLeuAlaValAlaProGlySerValAspAspLys 900
      |||:::
25771 .TCGAACCTCGCGCACACCGCGCGCGCGCGGTGATCA 25819
      |||:::
900 eArgGlyLeuAlaSerLysLeuHisGlnGluLeuAlaLeuGlnTyr 916
      |||:::
25820 AGATGCTCTGGCG.....CTGGCCACGCGAGCTCGCGCACCTG 25863
      |||:::
917 ValValSerSerSerAlaValArg.....GluAl 926
      |||:::
25864 CACCGCTCGACGCGCTGTCAGATGATGGACGCGCGCGCGTGA 25913
      |||:::
926 aIleLeuGlnHisAla.....TyrPhePheGlnLeuMetValLys 941
      |||:::
25914 GTTCTGTGACGAGCGCAGCGCTGCTC..... 25941
      |||:::
941 eMetAlaLeuHisLeuLeuGlnGlyGlnArgLeuAspThrProArgLys 957
      |||:::
25942 .....CAGCGCGCGCGAGCGCGCGCGCG 25965
      |||:::
958 LeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGlySe 974
      |||:::
25966 rValGlyLeuGlnValIleThrArgValHisLysAspValGlyLeuAlaG 991
      |||:::
25980 GTTGGGATGACGCGC...ACCAACGCGCACCTGTATCGACGAGCGCG 26026
      |||:::
991 LuHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuSerLeu 1007
      |||:::
26027 CGAGCGCCACCGCG.....CCGACACTGCTC..... 26052
      |||:::
1008 ValAspArgGlyPheValPheSerLeuValArgAlaHisTyrLysGlnVa 1024
      |||:::
26052 ..... 26052
      |||:::
1024 AlaThrArgLeuGlnSerSerProAsnProAlaAlaLeuLeuThrLeu 1040
      |||:::
26053 .....CGCCCGACCGCGCGCGCGCGCGCGCGCTCT 26083
      |||:::
1041 .....ArgMetGluPheThrArgGlyLeuCysSerHisGlnHisTyrVa 1055
      |||:::
26084 GGTCCGAGAGTGTGGACAGAGTGAACCTGCCCTGATGATGTCGCG 26133
      |||:::
1055 LThrLeuAsnLeuProCysCys.....ProLeuSerProPro 1068
      |||:::
26134 CACAGAGAACCGCGCTGCGCGACGAGCGCGCGCGCTGCCGCGCACCT 26183
      |||:::
1068 LaSerProSerProSerValSerSerThrThrSerGlnSerSerThrPhe 1084
      |||:::
26184 GCTCGCGCACCGCGAGCTGACCGCGCGACGCTGCGCTACA..... 26224
      |||:::
1085 SerSerGlnAlaProAspProLysValThrSerMetPheGlnLeuSerG 1101
      |||:::
26225 CCTCATTCACACCGCGCGCGCG..... 26247
      |||:::
1101 yProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuLeuThrGlnLeuA 1118
      |||:::
26248 ....TCGAGACAGCGCGCGCGCTGCTCGGAGAACTTCACGAGGCTGA 26293
      |||:::
1118 LaLeuAlaLeuGluProGluAlaGlnGlyAlaPheLeuLeuHisLysLys 1134
      |||:::
26294 TCGCGCGCGCTCGACGACCTGTCGAAGGC..... 26322
      |||:::
1135 AlaIleSerAlaValHisSerLeuLeuCysGlyHisAspThrAspPro 1150
      |||:::
26323 .....CGACCGACCGCGCTGCTGCGGCGCGACCGCGCGCACCTC 26363
      |||:::
1151 ArgTyrAlaGlnAlaThrValLysAlaArgValAlaGlnLeuTyrLeuPr 1167
      |||:::
26364 CGA.....CCAGCTCGTGT...CGTCTCCCGCGCG 26392
      |||:::
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
      |||:::
26393 AGGCGCTCGACGTGCGCGCGAGATGCGCGACG.....GCTG 26427
      |||:::
1184 LuGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
      |||:::
26428 CTGGCCGCTCCAGCGCGCTCGGACTC.....CTT 26456
      |||:::
1201 GluGlyGlnGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217
      |||:::
26457 CTGGAGACCGCGCGCGCTGCGACCTCGCGCTCGCGCCACCTCGGCT 26506
      |||:::
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerGln 1234
      |||:::
26507 GGTCCGCTCTGGACGACTGCGCGCGGAAACCGCGCGCGCTCTCGCA. 26555
      |||:::
1234 LyrProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlnSerSer 1250
      |||:::
26556 .....CGGGTCCAGCTGTGCGACCGCTGCTGTCCACCATGAT 26594
      |||:::
1251 ArgThrLeuLeuAlaCysValLeuThrValLeuLysAsnThrGluProA 1267
      |||:::
26595 GGT.....CTGCTCGCGCGACAGCTGGCGTTCGCTGGCGCGCAACCG 26638
      |||:::
1267 LaLeuLeu.....GlnArgTyrAla 1273
      |||:::
26639 CCGCGGTCTCGTCACTCCACGCGCGAGATCGCGCGCGCTACGTGCTCC 26688
      |||:::
1274 ThrAspLeuThrLeuProGlnLeuGlyArgLeuAspLeu..... 1287
      |||:::
26689 GCGCGCTGACGCTGTGACAGCGCGCGCAGTCGCTGCGCGACGCA 26738
      |||:::
1288 .LeuTyrLeuCysLeuAlaAlaPheGlnTyrLysGlyLysAlaPheG 1304
      |||:::
26739 GCGTGTGCTGCGCGCTGCGC.....GCCAAGGCGCGCATGG 26773
      |||:::
1304 LuArgIleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArg 1320
      |||:::
26774 TCGCGCGTAC.....CTGTCCGAACGACCTGCTGCCGCG 26811
      |||:::
1321 LeuGlnGlu..... 1323
      |||:::
26812 CTGGAGCGCTGGAGCGACCGCGCTGCGCGCGCGCTCAACGCGCGCA 26861
      |||:::
1324 .....AlaIleLeuGlyThrIleGlyAlaArgGlnGlnMetValArg. 1337
      |||:::
26862 GACCTGCGCGCTTCGCGGAGACCGGAGCGCGCTGCGGAGTGTGCCG 26911
      |||:::
1338 ArgSerArgGluArg.....SerProGlnLysAsnPr 1348
      |||:::
26912 AACCTGGGTGCGAGGCGCTGCGACCGCGCGCGCGCGCGGTGAGACAC 26961
      |||:::
1348 oGlnAsnValArgTyrArgLysSerValThrHisTyrGlnThrSerA 1365
      |||:::
26962 GCGCGGACCTGCGCGCGAGCTG.....ACAGCTGGAAGCGCGACCTGCG 27005
      |||:::
1365 sParGValAspLys...ThrLysAspGlnMetGlnHisGlnAlaLeuVa 1380
      |||:::
27006 GAAGTGTCTGCGCGCGCGCGCGCGCGCACTGCAATCCGCTTCACT 27055
      |||:::
1380 LglnGlyAsnLeuAlaThrGlu 1387
      |||:::
27056 CGAGGTCACCGGAGACTGAT 27077
      |||:::
seq_name: /cgn2_6/prodata/2/lna/5B_comb.seq:us-08-804-198-1
seq_documentation_block:
```

```
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kubstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470
; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14045..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1

Alignment scores:
  Quality: 182.00      Length: 1649
  Ratio: 0.285        Gaps: 89
  Percent Similarity: 38.690  Percent Identity: 21.953

Alignment_block:
US-09-736-968A-2 x US-08-804-198-1 ..
Align seg 1/1 to: US-08-804-198-1 from: 1 to: 44377

2 AaalaSerGluArgAlaPheAla.HisLysIleAsnArgThrVala 18
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
23189 GCCTGCTCGACGACGATCAACGCGACCGCGCTCACCCTCCTC 23238
18 laalagluValArgLysglInValSerArgGluArgSerGlySerProHis 34
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
23239 TCGCT.....GCTGGACTGTGCACCGACGCGAGCCGCGACACCCGCG 23282
35 SerSerArgArgCysSerSerSerSerLeuGlyValProLeuThrGluVala 51
:::|||||
23283 CGTGGCCGTGCG..... 23294
51 lGluProLeuAspPheGluAspValLeuLeuSerArgProProAspAlag 68
||| |||
23295 ..... 23316
68 lndProGlyProLeuArgAspLeuValGluPheProAlaAspAspLeuGlu 84
:::||||| |||
23317 GTCCAGGCCCTGGCCGA..... 23333
85 LeuLeuLeuGlnProArgGluCysArgThrThrGluProGlyIleProIly 101
|||||:|||||:|||||:|||||:|||||:
23334 .....CCTGGCAGCAGCAGCAGCGCT..... 23354
101 sAspGluLysLeuAspAlaGlnValArgAlaValaGluMetTyrIleG 118
:::|||||:|||||:|||||:|||||:|||||:
23355 .....GTGGACGTCACCTGGCGCGGT.....CGCCACCGCC 23388
118 lAspThrValIleValHisArgArg..... 126
23389 CCGGACGAACCTGCCGTGCACCGCGGTGCCAGCTGTGGGGCTGGGGCGG 23438
127 .....TyrGlnTyrIle 130
23439 GGTGGCCGCGCTGAGCTCCGAGGTGTGGGGCGGCTTCACACGACCTTC 23488
130 useAlaAlaTyrSerProValThrThrAspThrGlnArgGluGlnL 147
|||||:|||||:|||||:|||||:|||||:
23489 CCGCGGGCCCGACGCCCGGCTCTGGACCGCTCCGCGGT..... 23531
147 ySgLYLeuProArgGlnValPheGluGlnAspAlaSerGlyAspGluArg 163
|||||:|||||:|||||:|||||:|||||:
23532 .....CCTGCGCGACCC.....CGGCGG 23549
164 SerGlyProGluAspSerAsnAspSerArgArg.....Glyse 176
|||||:|||||:|||||:|||||:|||||:
23550 CGAGGACGATCGCCGTCGATGCGGGGCGCTTCGGCCGCGCGGTCC 23599
176 rGlySerProGluAspThrProArgSerSerGlyAlaSerIlePheAla 193
:::||||| |||
23600 TGGGGAACCGCGCGACTCCGCGCCCGCC..... 23627
193 sPLeuArgAsnLeuAlaAlaAspSerLeuLeuProSerLeuLeuGlnArg 209
:::|||||:|||||:|||||:|||||:|||||:
23628 .....GGCTGGCGCGCGCGCGGACCGCTCT.....CATCGC 23660
210 AlaAlaProGluAsp..... 214
|||||:|||||
23661 CGGCGACTCGACGAGGTGCCCGCGCGACTGTGTCGCTCCTCTCGAGG 23710
215 .....ValAspA 217
23711 ACGGCGCGGACCGCGGTGCTGTGGCGGACCGACCGCCCGCGGACGCG 23760
217 rGArgAsnGluThrLeuArgArgGlnHisArgProProAlaLeuLeuThr 233
|||||:|||||:|||||:|||||:|||||:|||||:
23761 GCGCGCCCGCGACTGACCGGCTCTCCCTGTCGCCGCTGGCGTGCAGCT 23810
234 LeuTyrPro.....AlaProAspGluAspGluAlaValG 245
||| ||| |||
23811 CACCGACCGCGCGCGACTGCGCGCTGCTCGACGACGACCGCCCGACCG 23860
245 uArgCysSerArgProGluPro..... 252
:::|||||:|||||:|||||:|||||:|||||:
23861 TCGCGGTGACGCGCGCGCTGTGTCGCGCTGCGCGCGCTGCGGAGAGG 23910
253 ..ProArgGluHisPheGlyGlnArgIleLeuValLysCysLeuSerIle 268
|||||:|||||:|||||:|||||:|||||:|||||:
23911 GCACCGCGCGACAT...CGCGCGCGCGCTGCGCGCGACGACGCGCGCG 23957
```

```

268 uLysPheGluIleGluProIlePheGlyIleLeuAlaLeuTyrA 285
      :::::::::: :::: ::::: :::::
23958 CGGCCACCTGTCGACCTGGCGCGCGGCTCGACCGCGT... 24003
      :::::::::: :::: ::::: :::::
285 sPValArgGluLysLysIleSerGluAsnPheTyrPheAspLeuAsn 301
      :::::::::: :::: ::::: :::::
24004 .....GTGCTGTTCTCTCGCTCGCTCGGAGTGTGG 24033
      :::::::::: :::: ::::: :::::
302 SerAspSerMetLysGlyLeuLeuArgAlaHisGlyThrHis...ProAl 317
      :::::::::: :::: ::::: :::::
24034 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCGACG 24083
      :::::::::: :::: ::::: :::::
317 aIleSerThrIleuAlaArgSerAla.....IlePheSerValThrT 331
      :::::::::: :::: ::::: :::::
24084 GCTGGCGACAGCGCGCGCGCGCGCGCGCGCGCTTCCTCGCGCGCT 24133
      :::::::::: :::: ::::: :::::
331 YrPro..... 332
      :::::::::: :::: ::::: :::::
24134 GGAGCCCTGGGGCGGAGGACGCCCGCGCGCGCGCGAGCGAGCTTC 24183
      :::::::::: :::: ::::: :::::
333 .....SerProAspIlePheLeuValI 340
      :::::::::: :::: ::::: :::::
24184 CTCAGCCCGCGCGCGCGCTGCTCCCTCGACCCGAC...CAGCGGCTGG 24230
      :::::::::: :::: ::::: :::::
340 eLysLeuGluLysValLeuGlnGlnGlyAspIleSerGluCys..Cys.. 355
      :::::::::: :::: ::::: :::::
24231 GACCTGCGCGCATGCTGGAGCGCGC....AGCGCTGCGGTGCGG 24274
      :::::::::: :::: ::::: :::::
356 .....GluProTyrMetValLeuLysGluValAspIle 367
      :::::::::: :::: ::::: :::::
24275 TCGCGGAGCTGAGTGGAGCGCGCTTCGCGCTCTACACCTGGGTGCT 24324
      :::::::::: :::: ::::: :::::
367 IAluYAsnLysGluLysLeuGluLysLeuArgLeuAlaIleGluIlePhe 383
      :::::::::: :::: ::::: :::::
24325 CCGCGCGCTCTTCGACGACATCCGCGAGTGGAGCGCGCTCGCGCGG 24374
      :::::::::: :::: ::::: :::::
384 CysThrArgLeuGlyArgTyrArgMetProPheAlaThrPheAlaValH 400
      :::::::::: :::: ::::: :::::
24375 CGAACTGCGCGCGGACCGGAGACTGAC.....CA 24406
      :::::::::: :::: ::::: :::::
400 sLeuAlaAsnIleValSerSerAlaGlyIleAspArgAspSerAsps 417
      :::::::::: :::: ::::: :::::
24407 CCTC.....CGAACTGTCGCGGAGCTGACGCGCGGAGTCCGCG 24444
      :::::::::: :::: ::::: :::::
417 eArgLysGluArgArgProAlaThrPheAspArgArgGlyPro 433
      :::::::::: :::: ::::: :::::
24445 CACAAGCGGACGCGCACCTGCTGCGGTGCGGCGCACACGCGCGCGC 24494
      :::::::::: :::: ::::: :::::
434 GlnAspArgAla..SerSerGlyAspAspAlaCysSerPheSerGlyPhe 450
      :::::::::: :::: ::::: :::::
24495 CGTCTCGGACAGTCTCGGCGGAGGAGGAGGAGCGCGCGCGCTTCC 24544
      :::::::::: :::: ::::: :::::
450 rGrProAla.....ThrLeuThrValThrAsnPhePheLysGlnIle 463
      :::::::::: :::: ::::: :::::
24545 GCGACCTCGGCTTCGACTCGCTGACGCGCTCGAAGT... 24582
      :::::::::: :::: ::::: :::::
464 ALGLuArgLeuSerAspGluAspLeuPheLys.....PheLe 476
      :::::::::: :::: ::::: :::::
24583 CCGGACCGGCTCA6CACGACCGGCGCTCAAACTGCCACCTCCCTGTG 24632
      :::::::::: :::: ::::: :::::
476 uAlaAspMetArgArgProSerSerLeuLeuArgArgLeu..... 489
      :::::::::: :::: ::::: :::::
24633 CTTCGACCACTCCAGCGCGCGCTCGCGCGCGCGCTGGTGAAGAAC 24682
      :::::::::: :::: ::::: :::::
490 .....Arg..ProValThrAlaGlnLeuIle 497
      :::::::::: :::: ::::: :::::
24683 TCCCTCGCGCGAAGACACCGCGGAGCGCGCGCGCGACACCGCGGTA 24732
      :::::::::: :::: ::::: :::::
497 sIleAspIleSerProAlaProGluAsn.....ProH 508
      :::::::::: :::: ::::: :::::
24733 CGGACGAGAGCGCCATCGCATCATGCGCATGGCTCGCGGTGCGCGG 24782
      :::::::::: :::: ::::: :::::

```

```

508 tSPheCysLeuSerProGluLeuLeu.....HisIleLysPro 520
      :::::::::: :::: ::::: :::::
24783 CGGGGTGCACTCCCGAGGACTGTGGGACTGTGACCGGTGGAGCG 24832
      :::::::::: :::: ::::: :::::
521 TyrProAspProArgGlyArgProThrLysGluIleLeuGluIlePheAl 537
      :::::::::: :::: ::::: :::::
24833 ACGCCATCACCCCTTCGCGACCAACCGGGAGTGGAGCAAGACCTTC 24882
      :::::::::: :::: ::::: :::::
537 a.ArgGlu.....ValTyrAlaProHisThrSerTyrAlaAsn 549
      :::::::::: :::: ::::: :::::
24883 TACGACCCCGACCCGACCTCGCGGCGACCACTACCTAGTGGCGCAGG 24932
      :::::::::: :::: ::::: :::::
550 LeuLeuTyrValTyrProHisSerLeuAsnPheSerSerArgGlnLys 566
      :::::::::: :::: ::::: :::::
24933 .....CGGTTCTCGACAGACGCGCGG 24955
      :::::::::: :::: ::::: :::::
566 rValArgAsnLeuAlaValArgValGln.....TyrMetThrGlyGlu 580
      :::::::::: :::: ::::: :::::
24956 AGTTGACACCCCGCTTCGCGCATCATGCCCGCGAGCGCTGGCATG 25005
      :::::::::: :::: ::::: :::::
581 AspProSerGlnAlaLeuProValIlePheGlyLysSerSerCysSerG 597
      :::::::::: :::: ::::: :::::
25006 GACCCGACGAGCGGCTGATCTG.....GACGCTCTGGGAGTC 25046
      :::::::::: :::: ::::: :::::
597 uPheThrArgGluAlaPheThrProValValTyrHisAsnLysSerProG 614
      :::::::::: :::: ::::: :::::
25047 CTTCGAACGCGCGCGCATCGACCCGCTCGACTGCGCGCGCAC... 25089
      :::::::::: :::: ::::: :::::
614 IuPheTyrGluGluuPheLysLeuHisLeuProAlaCysValThrGluAsn 630
      :::::::::: :::: ::::: :::::
25089 ..... 25089
      :::::::::: :::: ::::: :::::
631 HisHisLeuLeuPheThrPheThrHisValSerCysGlnProArgProG 647
      :::::::::: :::: ::::: :::::
25090 .....CGACCGG 25097
      :::::::::: :::: ::::: :::::
647 yThrAlaLeuGluThrProValGlyPheThrThrIleProLeuGlnIle 663
      :::::::::: :::: ::::: :::::
25098 GGTCTTCGTGCGCACCAAC...GGACAGCATACGTGCGGCTCTCCAGG 25144
      :::::::::: :::: ::::: :::::
664 .....HisGlyArgLeuArgThrGlyProPheCysLeu 674
      :::::::::: :::: ::::: :::::
25145 ACGGCGACGAGACTTCGACGCGCTACATCGCCGCGG... 25182
      :::::::::: :::: ::::: :::::
675 ProValSerValAspGlnProProProSerTyrSerValLeuThrProAs 691
      :::::::::: :::: ::::: :::::
25183 .....AACTCGGACGCGGAGTCCGCGGCG 25208
      :::::::::: :::: ::::: :::::
691 rValAlaLeuProGlyMetArgTyrValAspGlyHisLysGlyValPheS 708
      :::::::::: :::: ::::: :::::
25209 GCTTCCTTAC.....GTCTTCG 25225
      :::::::::: :::: ::::: :::::
708 eValGluLeuThrAlaValSerSerValHisProGlnAspProTyrLeu 724
      :::::::::: :::: ::::: :::::
25226 GACTGGAGGCGCGCGCTCAC... 25248
      :::::::::: :::: ::::: :::::
725 AspLysPhePheThrLeuValHisValLeuGluGluGlyAlaPheProP 741
      :::::::::: :::: ::::: :::::
25248 ..... 25248
      :::::::::: :::: ::::: :::::
741 eArgLeuLysAspThrValLeuSerGluGlyAsnValGluGlnIleuAla 758
      :::::::::: :::: ::::: :::::
25249 .....GTGACACCGCTGCT 25264
      :::::::::: :::: ::::: :::::
758 rGAlaSerLeuAlaLeuArgLeuAla..... 767
      :::::::::: :::: ::::: :::::
25265 CCGCTCTCTGCGGACCTGACCTGGCGGTGCACTGCTGCGCGCGGCG 25314
      :::::::::: :::: ::::: :::::
768 .....SerProG 770
      :::::::::: :::: ::::: :::::
25315 GAATGCGACTAGCGCTCGCGGCGGCGCACGCTGATGTCCACCCGGA 25364
      :::::::::: :::: ::::: :::::
770 uProLeuValAlaPheSerHisHisValLeuAspLysLeuValArgLeu 787
      :::::::::: :::: ::::: :::::

```

```
25365 GATGCTGTGTGGAGTCCGCCGT.....CAGCGAG 25393
      ||||| |||:.....
787 aAlIeArProProlIeIleSerGlyInIleValAsnLeuGlyArgIy 803
      ||| ||| :|||: |||
25394 CGGTGTGCGCGGAGGGCCGACAGCGGTCTGCGGAGCGCGCGAGCGG 25443
      ||| ||| :|||: |||
804 ALAPheGluNALaMetAlaHisValValSerLeuValHisArgSerLeuI 820
      :||| :||| :|||: |||
25444 GTGCGTCTCCCGAGGAGCGGGGATGCTGTGTGAGCGGCTGTGCGA 25493
      :||| :||| :|||: |||
820 uAlaAlaGlnAspAlaArgGlyHisCysProGlnLeuAlaAlaTyValH 837
      |||| :||| ||| ||| ||| |||
25494 GCGCGACAG.....AAGGCCAT...CCGTACTGGCGGTGTGTCGGG 25534
      ||||| :||| ||| ||| ||| |||
837 lTyAlaAPheArGlyLeuProGly.....ThrGluProSer 848
      ||| :||| ||| ||| ||| |||
25535 GCAGTGCCTGCACACAGACGCTGCCAGCACGGCTCACCCGACCCAGC 25584
      ||| :||| ||| ||| ||| |||
849 LeuPro.....AspGlyAl 853
      ||| :||| ||| ||| ||| |||
25585 GGGCCGCCAGAGCGGGGTGATACGGAGCGGCTGGCCAGCGGGGCT 25634
      ||| :||| ||| ||| ||| |||
853 aProProValThTyAlaGlnAlaAlaThrLeuAlaArgGlySerGlyArgP 870
      ||||| :||| ||| ||| ||| |||
25635 GACGCCCGCGACGTGGACGGGTGAG...GGCAGCGACACCGGACGCG 25681
      ||||| :||| ||| ||| ||| |||
870 ro..... 870
      ||| :||| ||| ||| ||| |||
25682 CGCTCGCGGACCCCATCGAGCGCGGCGCTGTCGCCACGTACGGCGG 25731
      ||| :||| ||| ||| ||| |||
871 .....AlaSerLeuTyLeuAlaArgSerLysSerIleSerse 883
      ||| :||| ||| ||| ||| |||
25732 GACCGCGCGACGCCCGCTGTGTGGCTGTGCTGTGAG..... 25770
      ||| :||| ||| ||| ||| |||
883 rSerAsnProAspLeuAlaValAlaAlaProGlySerValAspAspGluValS 900
      ||||| :||| ||| ||| ||| |||
25771 .TGGAAATGCGGCGACACCCAGCGCGCGCGGTGTGCGCGGGGTATTA 25819
      ||||| :||| ||| ||| ||| |||
900 eArArgIleLeuAlaSerLysLeuLeuHisGluLeuAlaLeuGlnTrpP 916
      :||| :||| ||| ||| ||| |||
25820 AGATGTGTGTGGCG.....CTGGCGCACGGAGCTGCCGCGCACCTGTG 25863
      ||||| :||| ||| ||| ||| |||
917 ValValSerSerSerAlaValArg.....GluAl 926
      ||||| :||| ||| ||| ||| |||
25864 CACCGGTGACGCGGCTGTCCAGATGTGGAGCGCGGCGCGCTGGA 25913
      ||||| :||| ||| ||| ||| |||
926 aIleLeuGlnHisAla.....TrpPhePheGlnLeuMetValLysS 941
      :||| :||| ||| ||| ||| |||
25914 GTTGTGTGAGGAGCGCGCGCTGTGCTC..... 25941
      ||||| :||| ||| ||| ||| |||
941 eMetAlaLeuHisLeuLeuLeuGlyGlnArgLeuAspThrProArgLys 957
      ||||| :||| ||| ||| ||| |||
25942 .....CAGCGGCGCGAGGGGCGCGCGCGG 25965
      ||||| :||| ||| ||| ||| |||
958 LeuArgPheProGlyArgPheLeuAspAspIleThrAlaLeuValGlyse 974
      ||||| :||| ||| ||| ||| |||
25966 .....GCGGCGATCTCTC 25979
      ||||| :||| ||| ||| ||| |||
974 rValGlyLeuGluValIleThrArgValHisLysAspValGluLeuAlaG 991
      ||||| :||| ||| ||| ||| |||
25980 GTTGGCATCTCAGCGGC...ACCAACGCGCACCTGTGATCAGGAGACCCG 26026
      ||||| :||| ||| ||| ||| |||
991 LuHisLeuAsnAlaSerLeuAlaPhePheLeuSerAspLeuLeuSerLeu 1007
      :||| :||| ||| ||| ||| |||
26027 CGGAGCCACCGCG.....CCGCAACTGCTC..... 26052
      ||||| :||| ||| ||| ||| |||
1008 ValAspArgGlyPheValPheSerLeuValArgAlaHisTyryGlnVal 1024
      :||| :||| ||| ||| ||| |||
26052 ..... 26052
      ||||| :||| ||| ||| ||| |||
1024 lAlaThrArgLeuGlnSerSerProAsnProAlaAla.LeuLeuThrLeu 1040
      :||| :||| ||| ||| ||| |||
```

```
26053 .....GCGCGCGAACCGCGCGCCGACGCGGCGAGCTCT 26083
      ||||| :||| ||| ||| ||| |||
1041 .....ArgMetGluPheThrArgIleLeuCysSerHisGlnHisTyryVa 1055
      ||||| :||| ||| ||| ||| |||
26084 GGTTCGAGGAGTGTGTGACAGAGTGAACCTGCCCTGATGATGTCTCCG 26133
      ||||| :||| ||| ||| ||| |||
1055 lThrLeuAsnLeuProCysCys.....ProLeuSerProProA 1068
      ||||| :||| ||| ||| ||| |||
26134 CACACAGACGCGCCTGTGCGACGACGCGCGCGCGCGCTGCGCGGACCT 26183
      ||||| :||| ||| ||| ||| |||
1068 lAserProSerProSerValSerSerThrThrSerGlnSerSerThrPhe 1084
      :||| :||| ||| ||| ||| |||
26184 GCTGCGCCACCGCAGCTGACCCGCGCAGAGCTGCGCTACA..... 26224
      ||||| :||| ||| ||| ||| |||
1085 SerSerGlnAlaProAspProLysValThrSerMetPheGluLeuSerG 1101
      ||||| :||| ||| ||| ||| |||
26225 CCTCATACACCGACCGACCGG..... 26247
      ||||| :||| ||| ||| ||| |||
1101 yProPheArgGlnGlnHisPheLeuAlaGlyLeuLeuThrGluLeuAl 1118
      ||||| :||| ||| ||| ||| |||
26248 ....TTCGACACGCGGCGCGCGCTGCGGCGAGAACTTCACGAGCTGA 26293
      ||||| :||| ||| ||| ||| |||
1118 lAlaLeuAlaLeuGluProGluAlaGluGlyAlaPheLeuLeuHisLys 1134
      ||||| :||| ||| ||| ||| |||
26294 TCGCGGCTCTGACGACCTGTGTGAAGG..... 26322
      ||||| :||| ||| ||| ||| |||
1135 AlAlIleSerAlaValHisSerLeuLeu.Cys.GlyHisAspThrAspPro 1150
      ||||| :||| ||| ||| ||| |||
26323 .....CGACCGACCGCGCTGTGTGTGCGGCGACCGCGCACCTTC 26363
      ||||| :||| ||| ||| ||| |||
1151 ArgTyAlaGluAlaThrValLysAlaArgValAlaGluLeuTyLeuPr 1167
      ||||| :||| ||| ||| ||| |||
26364 CGA.....CCAGGTCTGT.....CGTCTTCCCGGCG 26392
      ||||| :||| ||| ||| ||| |||
1167 oLeuLeuSerIleAlaArgAspThrLeuProArgLeuHisAspPheAlaG 1184
      ||||| :||| ||| ||| ||| |||
26393 AGGCTCTGACGTGGCGCGAGATGCGCGAGC.....GCTG 26427
      ||||| :||| ||| ||| ||| |||
1184 LuGlyProGlyGlnArgSerArgLeuAlaSerMetLeuAspSerAspThr 1200
      ||||| :||| ||| ||| ||| |||
26428 CTGGCGCGCTCCACGCGCTCGGCTC.....CTT 26456
      ||||| :||| ||| ||| ||| |||
1201 GluGlyGluGlyAspIleAlaGlyThrIleAsnProSerValAlaMetAl 1217
      ||||| :||| ||| ||| ||| |||
26457 CMTGAGACCGCGCGCGCTGCGACCTCGCGCGCGCCGACCTCGCGGT 26506
      ||||| :||| ||| ||| ||| |||
1217 aIleAlaGlyGlyProLeuAlaProGlySerArgAlaSerIleSerGln 1234
      :||| :||| ||| ||| ||| |||
26507 GGTCTGCTCTGAGTACTGCGCGGGAACCGCGCGCGCTGCTGCA. 26555
      ||||| :||| ||| ||| ||| |||
1234 LyProProThrAlaSerArgAlaGlyCysAlaLeuSerAlaGlnSerSer 1250
      :||| :||| ||| ||| ||| |||
26556 .....CGGCTGACGTGTGTGACACCGCTGTCTGTACATGAT 26594
      ||||| :||| ||| ||| ||| |||
1251 ArgThrLeuLeuAla.CysValLeuTrpValLeuLysAsnThrGluProA 1267
      ||||| :||| ||| ||| ||| |||
26595 GGT.....CTGCTGCGCGAGAGTGTGCTGTGCGGTGTGGAACCGG 26638
      ||||| :||| ||| ||| ||| |||
1267 lAlaLeu.....GlnArgTrpAla 1273
      ||||| :||| ||| ||| ||| |||
26639 CCGCGGTCTGTGCTACTCCAGGCGGAGATCCGCGCGCTTACGTGCGC 26688
      ||||| :||| ||| ||| ||| |||
1274 ThrAspLeuThrLeuProGlnLeuGlyArgLeuLeuAspLeu..... 1287
      ||||| :||| ||| ||| ||| |||
26689 GCGCCTGTGAGCTGAGACGCGCGCGCGATGCTGCGCTGTGAGCGCA 26738
      ||||| :||| ||| ||| ||| |||
1288 .LeuTyLeuCysLeuAlaPheGluTyryLysGlyLysAlaPheG 1304
      :||| :||| ||| ||| ||| |||
26739 GCGTGTGTGCGGTGCG.....GGCAAGGCGGCGATGG 26773
      ||||| :||| ||| ||| ||| |||
1304 LuArgIleAsnSerLeuThrPheLysLysSerLeuAspMetLysAlaArg 1320
      :||| :||| ||| ||| ||| |||
26774 TCGCGGTGAC.....CTGTCCGACGCGACCTGCTGCCCGC 26811
      ||||| :||| ||| ||| ||| |||
```

```
1321 LeuGluGlu..... 1323
      |||||
26812 CTGGAGCCCTGGAGCGACCGGCTCGCCGTCACGCGCCCGA 26861
1324 .....AlaIleLeuGlyThrIleGlyAlaArgGlnGluMetValArg. 1337
      |||||
26862 GACCTGCGCGCGTCTCCGGGAGCCCGACGCCCTTGCGGAGCTGTCGCCG 26911
1338 ArgSerArgGluArg.....SerProPheGlyAsnPr 1348
      |||||
26912 AACTCGGTGGGAGGGGCTGCACGCCGCCCATCCCGCGCTCGACACC 26961
1348 oGluAsnValArgTPrArgLysSerValThrHisTrpLysGlnThrSerA 1365
      |||||
26962 GCCGGGCACACACCGCGAGTCG.....ACACGCTGAGAGCCCACTCTGCG 27005
1365 sPARgValAspLys...ThrLys.AsprGluMetGlnHisGluAlaLeuVa 1380
      |||||
27006 GAAGGTGCTGCGCCGCTCGCGCCCGCACTCCGACATCCGTTCTACT 27055
1380 lGluGlyAsnLeuAlaThrGlu 1387
      |||||
27056 CGACGTCACCGGAGGACTGAT 27077
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 5, 2002, 02:45:43 ; Search time 899.83 Seconds
(without alignments)
12158.051 Million cell updates/sec

Title: US-09-736-968a-1
Perfect score: 6372
Sequence: 1 gagcgagcgagccatgctgc.....aagcaaaaaaaaaaaaaa 6372

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues
Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
N.Geneseq_032802:*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6372	100.0	6372	22	AAH43851
2	2140	33.6	2148	22	AAH43850
3	2132.8	33.5	6828	22	AAH41934
4	1729.4	27.1	7215	22	AAH07373
5	1465.2	23.0	1652	22	AAH99572
6	1446.6	22.7	4143	22	AAH41911
7	1446.6	22.7	4144	22	AAH41912
8	1446.6	22.7	4200	22	AAH41953
9	1195.8	18.8	3173	22	AAH41952

10	1031.8	16.2	4027	22	AAH07382	Human DNA associat
11	867.4	13.6	2014	22	AAH07381	Human DNA associat
12	829.4	13.0	1273	21	AAH093501	Human secreted pro
13	468.6	7.4	586	21	AAH75563	Human OREF ORF118
14	463.4	7.3	981	23	AAH70846	DNA encoding novel
15	448	7.0	2427	22	AAH161331	Human polynucleoti
16	407.8	6.4	2433	22	AAH159545	Human polynucleoti
17	407.4	6.4	9389	23	AAH17736	Drosophila melanog
18	385.4	6.0	966	24	AAH87322	DNA encoding novel
19	356.6	5.6	1834	23	AAH562625	CDNA sequence #412
20	322.4	5.1	3899	22	AAH32675	Human CDNA encodin
21	284	4.5	300	21	AAH00898	Human colon cancer
22	266.8	4.2	5589	23	AAH17737	Drosophila melanog
23	255.2	4.0	6816	21	AAH74524	Human OREF ORF79 p
24	255.2	4.0	7506	22	AAH19118	Angiogenesis assoc
25	243.8	3.8	417	21	AAH74475	Human OREF ORF30 p
26	238.2	3.7	321	22	AAH43871	Human CLASP-7 intr
27	238.2	3.7	512	22	AAH59901	Human foetal liver
28	238.2	3.7	3023	22	AAH17551	Human CDNA sequenc
29	237.6	3.7	4420	22	AAH52954	Human polynucleoti
30	232	3.6	232	22	AAH72442	Human foetal liver
31	225.4	3.5	5688	21	AAH87969	Human CLASP relate
32	225.4	3.5	5688	21	AAH14825	DNA encoding a hum
33	225.4	3.5	7277	24	AAH518951	Human CDNA encodin
34	216.6	3.4	281	22	AAH43857	Human CLASP-7 intr
35	213.8	3.4	1605	22	AAH07383	Human DNA associat
36	210	3.3	5214	21	AAH87968	Mouse CLASP-1 nucl
37	210	3.3	5214	21	AAH14824	DNA encoding a mur
38	199.8	3.1	2036	22	AAH14086	Human CDNA sequenc
39	194.8	3.1	2610	22	AAH32647	Human CDNA encodin
40	194.8	3.1	3472	22	AAH08358	Human CDNA encodin
41	191.6	3.0	4393	22	AAH08357	Human CDNA encodin
42	191.6	3.0	6454	22	AAH08334	Human polynucleoti
43	189.4	3.0	441	22	AAH84302	Human CLASP-7 intr
44	184.8	2.9	271	22	AAH43889	Human foetal liver
45	184.8	2.9	512	22	AAH61395	Human foetal liver

ALIGNMENTS

RESULT 1	AAH43851	AAH43851 standard; CDNA: 6372 BP.
ID	AAH43851	
XX	AAH43851:	
AC	04-SEP-2001 (first entry)	
XX		
DT		
XX		
DE	Human CLASP-7 encoding CDNA sequence SEQ ID NO:1.	
XX		
KW	Human; CLASP-7; cadherin-like asymmetry protein; immune response;	
KW	neuroprotective; antidiabetic; immunosuppressive; antihemetic;	
KW	antiarthritic; hypotensive; anti-HIV; cytostatic; immunostimulant;	
KW	antianaemic; antiinflammatory; ophthalmological; nephroprotectic;	
KW	antihypoid; antiasthmatic; antiallergic; antibacterial; gene therapy;	
KW	chromosome 19q13.2; autoimmune disease; multiple sclerosis; toxemia;	
KW	juvenile diabetes; rheumatoid arthritis; puritic ulcerarial papule;	
KW	hypertension; Rh incompatibility; ss.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	13..6156
FT		/*tag= a
XX		/product= "CLASP-7"
XX		
PN	W0200142295-A2.	
XX		
PD	14-JUN-2001.	
XX		
PF	13-DEC-2000; 2000WO-US34152.	
XX		

PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
PR 13-OCT-2000; 2000US-0240539.
PR 13-OCT-2000; 2000US-0240543.
XX

PA (ARBO-) ARBOR VITA CORP.

PI Lu P, Garman JD, Candia AF;

DR MPI; 2001-381641/40.

XX P-PSDB; AAB99541.

PT Novel cadherin-like asymmetry protein-7 and polynucleotides encoding
PT the polypeptide, useful for treating autoimmune disease,
PT hypersensitivity, preventing transplant rejection by modulating immune
PT response

XX Claim 1; Fig 5; 151pp; English.

XX The present invention describes a human cadherin-like asymmetry protein
XX (CLASP), designated CLASP-7. The CLASP-7 protein (I) and its encoding
XX nucleotide sequence (II) have activities including: neuroprotective;
XX antidiabetic; immunosuppressive; antirheumatic; antihypertensive; anti-HIV;
XX hypotensive; cytostatic; immunostimulant; antineoplastic; antiinflammatory;
XX ophthalmological; nephrotoxic; antithyroid; antitumoral; (III) are
XX antiallergic; and antibacterial. (II) and CLASP-7 antibodies (III) are
XX useful for detecting the CLASP-7 polypeptide. (II) is useful for
XX producing (I) by recombinant methods. (I) or its fragment are useful for
XX inhibiting an immune response in a cell such as T cell or B cell. A
XX pharmaceutical composition (C), comprising (I) or (II), can be useful
XX for treating CLASP-7-mediated disease such as an autoimmune disease
XX caused or exacerbated by increased activity of T helper cells. Autoimmune
XX diseases which can be treated using (C) include multiple sclerosis,
XX juvenile diabetes and rheumatoid arthritis. (I) is useful for treating
XX toxemia or pregnancy induced hypertension, pruritic urticarial papules
XX and Rh incompatibility. (I) is also useful as a diagnostic reagent for
XX immune and other disorders, since diseases characterized by
XX overproduction or depletion of lymphocytes in blood or other organs may
XX be detected by monitoring the level of (I) or its mRNA. CLASP-7 has been
XX mapped to the chromosomal location 19q13.2. The present sequence
XX encodes the human CLASP-7 protein as given in the present invention.

XX Sequence 6372 BP; 1273 A; 2050 C; 1849 G; 1200 T; 0 other;

Query Match 100.0%; Score 6372; DB 22; Length 6372;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 6372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gacgcgagacacatgctgctccctccgagcgccgcttcgcgcacacagatcaacagacg 60
DB 1 gacgcgagacacatgctgctccctccgagcgccgcttcgcgcacacagatcaacagacg 60
QY 61 gtcgcgcgacgagtgctgcgaagcaggtgtcccggaagcagtggtgtcccccacacg 120
DB 61 gtcgcgcgacgagtgctgcgaagcaggtgtcccggaagcagtggtgtcccccacacg 120
QY 121 aggcgcgtgcagcagctccctcgtgggttcccaactgactgaagtgtgaacccctgagatt 180
DB 121 aggcgcgtgcagcagctccctcgtgggttcccaactgactgaagtgtgaacccctgagatt 180
QY 181 gagagtgtaactctctgacccgcgcacacagatgctgagcccgggccctcaaggagactgtga 240
DB 181 gagagtgtaactctctgacccgcgcacacagatgctgagcccgggccctcaaggagactgtga 240

QY 241 gaattccagctgactgtgagctgctgtcgcagcccggaatcccgaccagag 300
DB 241 gaattccagctgactgtgagctgctgtcgcagcccggaatcccgaccagag 300
QY 301 cccgggattcccaagatgtaaaactgtgacccaggtgaaggcggtgtagatgtat 360
DB 301 cccgggattcccaagatgtaaaactgtgacccaggtgaaggcggtgtagatgtat 360
QY 361 attgagagctggtcatatgtccacaagaagatagtaacctgaatgacgataagcc 420
DB 361 attgagagctggtcatatgtccacaagaagatagtaacctgaatgacgataagcc 420
QY 421 gtcaacacagacacacagcgggagcgacagaaaggcctcccccgcacagttctgaacag 480
DB 421 gtcaacacagacacacagcgggagcgacagaaaggcctcccccgcacagttctgaacag 480
QY 481 gatgctcttgtagacgagaggtccggccctgaagatccgaatctccggcgtgtcgc 540
DB 481 gatgctcttgtagacgagaggtccggccctgaagatccgaatctccggcgtgtcgc 540
QY 541 ggtcccccgaagacacccctcggaagcagtggtgctcctctagcatctcgactgagagac 600
DB 541 ggtcccccgaagacacccctcggaagcagtggtgctcctctagcatctcgactgagagac 600
QY 601 ctgagcagctgacatctgtgcctctctctgtagagcgggcgccccaagatgtggac 660
DB 601 ctgagcagctgacatctgtgcctctctctgtagagcgggcgccccaagatgtggac 660
QY 661 cggcgcaatgaaaccttgtagcggcagacacggcccgccctgctcacccctacccg 720
DB 661 cggcgcaatgaaaccttgtagcggcagacacggcccgccctgctcacccctacccg 720
QY 721 gcaactgacgagtgaaacggtgaaacgctgtgacggccgacagacaccccgagac 780
DB 721 gcaactgacgagtgaaacggtgaaacgctgtgacggccgacagacaccccgagac 780
QY 781 ttgtgacaaagatctgtgtcaagtgtctgtcgtcaagtgtgaatgaaatgagcc 840
DB 781 ttgtgacaaagatctgtgtcaagtgtctgtcgtcaagtgtgaatgaaatgagcc 840
QY 841 attcttggaactggtcctgtatgatgtgcggagaagaagaagatctggagaacttc 900
DB 841 attcttggaactggtcctgtatgatgtgcggagaagaagaagatctggagaacttc 900
QY 901 tacttcgaactgaaactcgtgactccatgaaggggctgttcgggtcattgacacccact 960
DB 901 tacttcgaactgaaactcgtgactccatgaaggggctgttcgggtcattgacacccact 960
QY 961 gcaatctccacccctggccgctctgtccactctctctgtgacctaaccctacctaacc 1020
DB 961 gcaatctccacccctggccgctctgtccactctctctgtgacctaaccctacctaacc 1020
QY 1021 ttctgtgtcatcaagtgtgagaaggtgtcttcagcaagggaacatcagtgatgtctgtgag 1080
DB 1021 ttctgtgtcatcaagtgtgagaaggtgtcttcagcaagggaacatcagtgatgtctgtgag 1080
QY 1081 ccttacatgtgtgtgaaagaagtgtgacacagccaagaagaagaagaagctagaagaagctg 1140
DB 1081 ccttacatgtgtgtgaaagaagtgtgacacagccaagaagaagaagaagctagaagaagctg 1140
QY 1141 cgctctgacgacgagcagatctgtcacccgctgtggcgcgtacacgcatgtcccttcgcctg 1200
DB 1141 cgctctgacgacgagcagatctgtcacccgctgtggcgcgtacacgcatgtcccttcgcctg 1200
QY 1201 acggccgtgcaactgtgccaacatcgtggaagagcgctgggacgtggaacgggactctgac 1260
DB 1201 acggccgtgcaactgtgccaacatcgtggaagagcgctgggacgtggaacgggactctgac 1260
QY 1261 tcggagagcgagcgccggcagccttgaaacagacccgcgcttcggggcgcccaagacgg 1320
DB 1261 tcggagagcgagcgccggcagccttgaaacagacccgcgcttcggggcgcccaagacgg 1320
QY 1321 gcgagtagtgaggagcagacgctcagactctctgtgcttcctgtccagccaagctaatgttc 1380

Db 1321 gcsagatgagtgagacgacgcccgtccagctctctctgctccgaccgacgctaacgtctc 1380
QY 1381 acaaaactctttaaagcagagagctgagcgaactgaacgagagacctcttaagtctctg 1440
Db 1381 acaaaactctttaaagcagagagctgagcgaactgaacgagagacctcttaagtctctg 1440
QY 1441 gctgacatagagagcccgctcgctccctgctgctgagcgaactaagctctctgtgactgcccagctc 1500
Db 1441 gctgacatagagagcccgctcgctccctgctgctgagcgaactaagctctctgtgactgcccagctc 1500
QY 1501 aagatcgcgaattctctccgctctctgaaaatcccacattctgctctctccctgagatgctt 1560
Db 1501 aagatcgcgaattctctccgctctctgaaaatcccacattctgctctctccctgagatgctt 1560
QY 1561 catatcaagcccttaacccgagaccccgaggcgccgacccaagaagatctctgagatctccc 1620
Db 1561 catatcaagcccttaacccgagaccccgaggcgccgacccaagaagatctctgagatctccc 1620
QY 1621 gcccgagagctctatgccccccaataccagctaacagaaacctgtgtacgtgtacccgac 1680
Db 1621 gcccgagagctctatgccccccaataccagctaacagaaacctgtgtacgtgtacccgac 1680
QY 1681 agcctcaactttagcagacgcccagggctcccgctgctgagcgaacttgcgtgagtgagttac 1740
Db 1681 agcctcaactttagcagacgcccagggctcccgctgctgagcgaacttgcgtgagtgagttac 1740
QY 1741 atgacagcgcgagagaccccgacagagctctgcccgcgtcatcttctgcaagtccagcgtcag 1800
Db 1741 atgacagcgcgagagaccccgacagagctctgcccgcgtcatcttctgcaagtccagcgtcag 1800
QY 1801 gaatttaccgcgagagccttcaacccgctggtcttaccataaagaatcccccagatctaac 1860
Db 1801 gaatttaccgcgagagccttcaacccgctggtcttaccataaagaatcccccagatctaac 1860
QY 1861 gagagatcaagaactgacatcttccagctctgctgagacagaacacatcacactgtgtctaac 1920
Db 1861 gagagatcaagaactgacatcttccagctctgctgagacagaacacatcacactgtgtctaac 1920
QY 1921 ttctaacatgctcagctgcccagcccgcgagcgaactgcccctgagacacacccgttgagctt 1980
Db 1921 ttctaacatgctcagctgcccagcccgcgagcgaactgcccctgagacacacccgttgagctt 1980
QY 1981 acttgatctcccaactgctgtaagacagggcgctgaggaacccgccccttctgtctcccaatg 2040
Db 1981 acttgatctcccaactgctgtaagacagggcgctgaggaacccgccccttctgtctcccaatg 2040
QY 2041 tctgtgagacagccgcgcgcagctatctccgtgctcaaccccgatgctgagcttccgggc 2100
Db 2041 tctgtgagacagccgcgcgcagctatctccgtgctcaaccccgatgctgagcttccgggc 2100
QY 2101 atgcgctgggtgagcaggtctacaaagggcggtgttcaggtgtggaagctcaacgcccgtctct 2160
Db 2101 atgcgctgggtgagcaggtctacaaagggcggtgtgttcaggtgtggaagctcaacgcccgtctct 2160
QY 2161 gtagaacccccagagaccctactctgtaacaaatcttcaaccctgtgacagctccctgtaagag 2220
Db 2161 gtagaacccccagagaccctactctgtaacaaatcttcaaccctgtgacagctccctgtaagag 2220
QY 2221 gtagaccttccattcccggtcaaggaactgctgctgagcgaaggcaaaagtgtgagagag 2280
Db 2221 gtagaccttccattcccggtcaaggaactgctgctgagcgaaggcaaaagtgtgagagag 2280
QY 2281 ctgcggggcagctctgtgcaagcaatgctgctgagcgaagcccgaaacccctgtgtgctctcc 2340
Db 2281 ctgcggggcagctctgtgcaagcaatgctgctgagcgaagcccgaaacccctgtgtgctctcc 2340
QY 2341 caacacgtgctgtgaaagaactgctgctgctgtgtcatcagaagcccgatcatcagtgagcag 2400
Db 2341 caacacgtgctgtgaaagaactgctgctgctgtgtcatcagaagcccgatcatcagtgagcag 2400
QY 2401 attgtgaaccttgccgctgtgagaccttgaagcaatgagcccatgtgagcctgttacc 2460

Db 2401 attgtgaaccttgccgctgtgagaccttgaagcaatgagcccatgtgagcctgttacc 2460
QY 2461 cggagcccttgagagcagaccagatgccccggtcactgcccacagctgagctactagtc 2520
Db 2461 cggagcccttgagagcagaccagatgccccggtcactgcccacagctgagctactagtc 2520
QY 2521 cactaagccttccgctctctctgacatgagcccaagcctccggatgtgggccccctccagtg 2580
Db 2521 cactaagccttccgctctctctgacatgagcccaagcctccggatgtgggccccctccagtg 2580
QY 2581 acaatgagagctgccaacacgagcccggtgctctgtgtccgcccgaagcctctacagtcg 2640
Db 2581 acaatgagagctgccaacacgagcccggtgctctgtgtccgcccgaagcctctacagtcg 2640
QY 2641 cgttccaaagacatcaagcagacgaacaaactgaaactcgccgtgagccctggtctgtgata 2700
Db 2641 cgttccaaagacatcaagcagacgaacaaactgaaactcgccgtgagccctggtctgtgata 2700
QY 2701 gacgaagcttcccgacatcctgagccagaaagctgtcttcaagaaagatgctgtcagtg 2760
Db 2701 gacgaagcttcccgacatcctgagccagaaagctgtcttcaagaaagatgctgtcagtg 2760
QY 2761 gtgtgtcagcagcagtgccgtgacgagagcgaatctccagagacgctggtctctccag 2820
Db 2761 gtgtgtcagcagcagtgccgtgacgagagcgaatctccagagacgctggtctctccag 2820
QY 2821 ctcaatgtgaagatlatgagcctgtgacactgtgcttggccagagatagaacaccccgc 2880
Db 2821 ctcaatgtgaagatlatgagcctgtgacactgtgcttggccagagatagaacaccccgc 2880
QY 2881 aagctgcgcttcccccagagccttccctgagacataacatctctgtgtggctctgtgggc 2940
Db 2881 aagctgcgcttcccccagagccttccctgagacataacatctctgtgtggctctgtgggc 2940
QY 2941 ctggaagctcaacacccgtgtccaaagaagatgtgagctggtccgagacaccccaacgacag 3000
Db 2941 ctggaagctcaacacccgtgtccaaagaagatgtgagctggtccgagacaccccaacgacag 3000
QY 3001 ctggcttctctccctagtgaaactctgtccctggtgagacccgggcttctgtcttcaacgt 3060
Db 3001 ctggcttctctccctagtgaaactctgtccctggtgagacccgggcttctgtcttcaacgt 3060
QY 3061 gtcggggcccaactcaagcaggggtgagccagccgctccagctccgctcccttaaccagagcc 3120
Db 3061 gtcggggcccaactcaagcaggggtgagccagccgctccagctccgctcccttaaccagagcc 3120
QY 3121 ctgcttgacccctgacatgagatltcaaccgcgcatcctgtgacagcagacgactagtcgac 3180
Db 3121 ctgcttgacccctgacatgagatltcaaccgcgcatcctgtgacagcagacgactagtcgac 3180
QY 3181 ctcaacctccctgtgccccctgttcaactccagcctgcccctccctctgtctcc 3240
Db 3181 ctcaacctccctgtgccccctgttcaactccagcctgcccctccctctgtctcc 3240
QY 3241 accacctcccaagactccaacttctccagccaaagcccggaaccccaaggtgacacagatg 3300
Db 3241 accacctcccaagactccaacttctccagccaaagcccggaaccccaaggtgacacagatg 3300
QY 3301 ttccgaactgagtgaaacatctccgagcagacaaacttccatgctggtgctctcgtgagagag 3360
Db 3301 ttccgaactgagtgaaacatctccgagcagacaaacttccatgctggtgctctcgtgagagag 3360
QY 3361 ctggaacttgccctcgaacactgagcctgaaaggggcatcttctgtgtgacaaagagccatc 3420
Db 3361 ctggaacttgccctcgaacactgagcctgaaaggggcatcttctgtgtgacaaagagccatc 3420
QY 3421 agtgctgtgacagcctgtcatgtgtgccaatgaaactgaaaccccgctaacgcccagggccact 3480
Db 3421 agtgctgtgacagcctgtcatgtgtgccaatgaaactgaaaccccgctaacgcccagggccact 3480
QY 3481 gtagaagctgctgtgagcagagcctgtacactgcccactgcttccgatgtgacgggataactgt 3540
Db 3481 gtagaagctgctgtgagcagagcctgtacactgcccactgcttccgatgtgacgggataactgt 3540

QY 3541 ccaagctgcatgacttctgtgagggccaggtcagcgtgtcaagacttgccatagctt 3600
|||||
Db 3541 ccaaggtgcatgacttctgtgagggccaggtcagcgtgtcaagacttgccatagctt 3600
QY 3601 gactcagaacacagaagcggaaggacattgcggttaccatcaaacctctgtgscatg 3660
|||||
Db 3601 gactcagaacacagaagcggaaggacattgcggttaccatcaaacctctgtgscatg 3660
QY 3661 gccattctgtgccccccagccctctgctcccgggccagacatctcccaaggggccacaa 3720
|||||
Db 3661 gccattctgtgccccccagccctctgctcccgggccagacatctcccaaggggccacaa 3720
QY 3721 accggtctctcgcaaggctgtgtccctctctgtctgtagtcaagccggaacctgtgctgct 3780
|||||
Db 3721 accggtctctcgcaaggctgtgtccctctctgtctgtagtcaagccggaacctgtgctgct 3780
QY 3781 gtgtctgtggtgtgtctgtaaaaaaacacccgaagccggcgctccctgacagcgtcggtccactgtac 3840
3781 gtgtctgtggtgtgtctgtaaaaaaacacccgaagccggcgctccctgacagcgtcggtccactgtac 3840
QY 3841 acaactccccagctgggaacgtctgtctggaattgtgtacatttgctctgctgtctgtgag 3900
|||||
Db 3841 acaactccccagctgggaacgtctgtctggaattgtgtacatttgctctgctgtctgtgag 3900
QY 3901 tacagaagggaaaaaaggcctttgaaacgatalcaaacagcctcaacattcaaaaaatctctgata 3960
|||||
Db 3901 tacagaagggaaaaaaggcctttgaaacgatalcaaacagcctcaacattcaaaaaatctctgata 3960
QY 3961 atgaaggcggtgtgtagaagaagccattctgtgtgtacacatcggtgagctgtgcaagaagaatggtt 4020
|||||
Db 3961 atgaaggcggtgtgtagaagaagccattctgtgtgtacacatcggtgagctgtgcaagaagaatggtt 4020
QY 4021 cggcgaaagtctgttagaaggaaacccgttctggaatccagagaacgtgtgctcgcgcggaagagc 4080
|||||
Db 4021 cggcgaaagtctgttagaaggaaacccgttctggaatccagagaacgtgtgctcgcgcggaagagc 4080
QY 4081 gtcaacacactgtgaagcaaacctctcagaacccgtgtgacagaacaaagataaaaaatctgaaac 4140
4081 gtcaacacactgtgaagcaaacctctcagaacccgtgtgacagaacaaagataaaaaatctgaaac 4140
Db 4141 gaaagccttggtagaaggaaacccgtgtggaatccagagaacgtgtgctcgcgcggaagagc 4200
4141 gaaagccttggtagaaggaaacccgtgtggaatccagagaacgtgtgctcgcgcggaagagc 4200
QY 4201 gaagatacgtgtgacagacgtgtgattcttcagaagcccggaagacgttctctggtggcgagt 4260
|||||
Db 4201 gaagatacgtgtgacagacgtgtgattcttcagaagcccggaagacgttctctggtggcgagt 4260
QY 4261 ctgaagagttgtgtctgtacagacctgtggcaagtgtcccaagatgtccctctctctgacagatggc 4320
|||||
Db 4261 ctgaagagttgtgtctgtacagacctgtggcaagtgtcccaagatgtccctctctctgacagatggc 4320
QY 4321 ctggcgcccccaaggggcccttgtgtccaagtgtcccggaatctcgtgtctgtctgtagagagacag 4380
4321 ctggcgcccccaaggggcccttgtgtccaagtgtcccggaatctcgtgtgtctgtctgtagagagacag 4380
Db 4321 ctggcgcccccaaggggcccttgtgtccaagtgtcccggaatctcgtgtgtctgtctgtagagagacag 4380
QY 4381 gaagctgtgtgtccgaacctgtgctgaggtctctcagacatgtgtgacagccggtcagcc 4440
|||||
Db 4381 gaagctgtgtgtccgaacctgtgctgaggtctctcagacatgtgtgacagccggtcagcc 4440
QY 4441 atccgcagcagcagccagccgctctgctgtacctgtctcatgtggaacagaactctgagatctggc 4500
4441 atccgcagcagcagccagccgctctgctgtacctgtctcatgtggaacagaactctgagatctggc 4500
Db 4441 atccgcagcagcagccagccgctctgctgtacctgtctcatgtggaacagaactctgagatctggc 4500
QY 4501 cacaactttggcccggtgtgtagagatgacgtgtccacatgtctctctgtccctgtgtgtagagag 4560
4501 cacaactttggcccggtgtgtagagatgacgtgtccacatgtctctctgtgtccctgtgtgtagagag 4560
Db 4501 cacaactttggcccggtgtgtagagatgacgtgtccacatgtctctctgtgtccctgtgtgtagagag 4560
QY 4561 accgcagaactctagtgaaagacacctgtgacgttcaactaaaaacatctctacatctgt 4620
4561 accgcagaactctagtgaaagacacctgtgacgttcaactaaaaacatctctacatctgt 4620
Db 4561 accgcagaactctagtgaaagacacctgtgacgttcaactaaaaacatctctacatctgt 4620

QY 4621 gaggaggaatgtggtctgcggagacagccactctgcagaagcaggtccaggaactgtatctc 4680
|||||
Db 4621 gaggaggaatgtggtctgcggagacagccactctgcagaagcaggtccaggaactgtatctc 4680
QY 4681 aaactgtacatgtatctctgaacggacacgggtgaaagtgaagaaacacccaggaagacctgtag 4740
4681 aaactgtacatgtatctctgaacggacacgggtgaaagtgaagaaacacccaggaagacctgtag 4740
Db 4681 aaactgtacatgtatctctgaacggacacgggtgaaagtgaagaaacacccaggaagacctgtag 4740
QY 4741 atgtctcatcgacctctatgtatagaattgtcccgggggtctacccaggggtctaacgggaacctctcg 4800
4741 atgtctcatcgacctctatgtatagaattgtcccgggggtctacccaggggtctaacgggaacctctcg 4800
Db 4741 atgtctcatcgacctctatgtatagaattgtcccgggggtctacccaggggtctaacgggaacctctcg 4800
QY 4801 ctgacacttgtgtcagaaataatgtccgggaagcagccggagctgtgggtcacaacccagccgaagcc 4860
4801 ctgacacttgtgtcagaaataatgtccgggaagcagccggagctgtgggtcacaacccagccgaagcc 4860
Db 4801 ctgacacttgtgtcagaaataatgtccgggaagcagccggagctgtgggtcacaacccagccgaagcc 4860
QY 4861 gcccaagtgtcatgtgtgcaagcggcgccctctgtgtgagtgatccctcgccctgtctgtgaagac 4920
4861 gcccaagtgtcatgtgtgcaagcggcgccctctgtgtgagtgatccctcgccctgtctgtgaagac 4920
Db 4861 gcccaagtgtcatgtgtgcaagcggcgccctctgtgtgagtgatccctcgccctgtctgtgaagac 4920
QY 4921 cagcgccacccctggcccggtgtggtgtgtctctctccagaacatctcatccaaacgtgtcagag 4980
4921 cagcgccacccctggcccggtgtggtgtgtctctctccagaacatctcatccaaacgtgtcagag 4980
Db 4921 cagcgccacccctggcccggtgtggtgtgtctctctccagaacatctcatccaaacgtgtcagag 4980
QY 4981 gagtccgcacatctccgaagacatccctgtctgcgccgaagcgaaggaggtctctgtccgggaag 5040
4981 gagtccgcacatctccgaagacatccctgtctgcgccgaagcgaaggaggtctctgtccgggaag 5040
Db 4981 gagtccgcacatctccgaagacatccctgtctgcgccgaagcgaaggaggtctctgtccgggaag 5040
QY 5041 cacttcaatctgagctgtggggtctgtgaggtgtgtgtgtaaaacaggaagccggtctacttaccatg 5100
5041 cacttcaatctgagctgtggggtctgtgaggtgtgtgtgtaaaacaggaagccggtctacttaccatg 5100
Db 5041 cacttcaatctgagctgtggggtctgtgaggtgtgtgtgtaaaacaggaagccggtctacttaccatg 5100
QY 5101 ggcgggctctacagagcgggtgaatgaaggtctcaacaaagaaaccccatcccatctgtgaagcc 5160
5101 ggcgggctctacagagcgggtgaatgaaggtctcaacaaagaaaccccatcccatctgtgaagcc 5160
Db 5101 ggcgggctctacagagcgggtgaatgaaggtctcaacaaagaaaccccatcccatctgtgaagcc 5160
QY 5161 caacgtgtactcaagaagaagctgtgcggcggtgtgcagcgcaaacctgtcagagaggtccttaccacaa 5220
5161 caacgtgtactcaagaagaagctgtgcggcggtgtgcagcgcaaacctgtcagagaggtccttaccacaa 5220
Db 5161 caacgtgtactcaagaagaagctgtgcggcggtgtgcagcgcaaacctgtcagagaggtccttaccacaa 5220
QY 5221 atcatgtacacacagatctccgggtctgggaagcgggtgtctgggaaggtatttcgcggtgtggtctc 5280
5221 atcatgtacacacagatctccgggtctgggaagcgggtgtgtctgggaaggtatttcgcggtgtggtctc 5280
Db 5221 atcatgtacacacagatctccgggtctgggaagcgggtgtgtctgggaaggtatttcgcggtgtggtctc 5280
QY 5281 taacgggccacactctcggtgacactgtgatatgaagcagagttgtgtacaaaggagccatccatc 5340
5281 taacgggccacactctcggtgacactgtgatatgaagcagagttgtgtacaaaggagccatccatc 5340
Db 5281 taacgggccacactctcggtgacactgtgatatgaagcagagttgtgtacaaaggagccatccatc 5340
QY 5341 acgaagctgtgcagagatctcaacacgggtctggaagagttctacacggagagaattgtgcgac 5400
5341 acgaagctgtgcagagatctcaacacgggtctggaagagttctacacggagagaattgtgtgcgac 5400
Db 5341 acgaagctgtgcagagatctcaacacgggtctggaagagttctacacggagagaattgtgtgcgac 5400
QY 5401 gaagtcgtgtgagattatcaaaagactcttaacctgtggaacaaagtccaaagctgtgactcaaa 5460
5401 gaagtcgtgtgagattatcaaaagactcttaacctgtggaacaaagtccaaagctgtgactcaaa 5460
Db 5401 gaagtcgtgtgagattatcaaaagactcttaacctgtggaacaaagtccaaagctgtgactcaaa 5460
QY 5461 aagggctacatccagatatacgttatgtgaaacgttacttatacctaagctcaagac 5520
5461 aagggctacatccagatatacgttatgtgaaacgttacttatacctaagctcaagac 5520
Db 5461 aagggctacatccagatatacgttatgtgaaacgttacttatacctaagctcaagac 5520
QY 5521 cgggtgtgacctacttctgacggcaactatagggcgttcgcacatctcgtgtctgtcacgcccgttc 5580
5521 cgggtgtgacctacttctgacggcaactatagggcgttcgcacatctcgtgtctgtcacgcccgttc 5580
Db 5521 cgggtgtgacctacttctgacggcaactatagggcgttcgcacatctcgtgtctgtcacgcccgttc 5580
QY 5581 acgcgcgatatgggcgcgcacacgggggaactgtcccggaacaaacaaagcgttaagaacgtgtctc 5640
5581 acgcgcgatatgggcgcgcacacgggggaactgtcccggaacaaacaaagcgttaagaacgtgtctc 5640
Db 5581 acgcgcgatatgggcgcgcacacgggggaactgtcccggaacaaacaaagcgttaagaacgtgtctc 5640
QY 5641 agcacgcacacagccttcccttatacataaagactgtgataccgtgtgtgtgtgtccacgggagag 5700
5641 agcacgcacacagccttcccttatacataaagactgtgataccgtgtgtgtgtgtgtccacgggagag 5700
Db 5641 agcacgcacacagccttcccttatacataaagactgtgataccgtgtgtgtgtgtgtccacgggagag 5700
QY 5701 acggtgtgtacagcaggtgtgaggtgtgcatctgaggagatgtcagaagaagacacgggagctg 5760

Db 61 ggcagtgccagagtgccctctcttgcaagatgacctgcccacccagaggcccttg 120
QY 4345 tccaagttcccgagctgctgtctcgagagagacagagagctgtgtcccgacctgtgctc 4404
Db 121 tccaagttcccgagctgctgtctcgagagagacagagagctgtgtcccgacctgtgctc 180
QY 4405 aggtctctacagacactgtgagcagccatcagcacccatcccgacagccagcgctc 4464
Db 181 aggtctctacagacactgtgagcagccatcagcacccatcccgacagcgctc 240
QY 4465 ctgtatcctgtccatgagcagagaaacttcgagatcgccacaacttgcgcgtgtgaagatg 4524
Db 241 ctgtatcctgtccatgagcagagaaacttcgagatcgccacaacttgcgcgtgtgaagatg 300
QY 4525 caggtcaccatgtctctctgtccctgtgtgagcagcagcagaaacttcagtgaaagacac 4584
Db 301 caggtcaccatgtctctctgtccctgtgtgagcagcagcagaaacttcagtgaaagacac 360
QY 4585 ctgagcagttcatcctcaaaaacatctcaactatgtctgagagagacatgagcgctgagagac 4644
Db 361 ctgagcagttcatcctcaaaaacatctcaactatgtctgagagagacatgagcgctgagagac 420
QY 4645 agcaccttcgagagcgaggttcagagacctgtatgttcaactgtcacatgacctgagcgagac 4704
Db 421 agcaccttcgagagcgaggttcagagacctgtatgttcaactgtcacatgacctgagcgagac 480
QY 4705 acggtgaagatgaagagacacacacagagagacctgtatgttcaactgtcacatgacctgagac 4764
Db 481 acggtgaagatgaagagacacacacagagagacctgtatgttcaactgtcacatgacctgagac 540
QY 4765 atgtcccgagggtctacagaggtctcacaggaaccttcgctgtgacctgtgtgacgaacatg 4824
Db 541 atgtcccgagggtctacagaggtctcacaggaaccttcgctgtgacctgtgtgacgaacatg 600
QY 4825 gggagagcagcgagcggtgtggggaacacagcgccagagcgccccaagtgcagtgtgagcgcgcc 4884
Db 601 gggagagcagcgagcggtgtggggaacacagcgccagagcgccccaagtgcagtgtgagcgcgcc 660
QY 4885 gccctctgtgctgtgtaacctgcgcctgtctcgagagacagcgccacccctgcgcgtgtgctgc 4944
Db 661 gccctctgtgctgtgtaacctgcgcctgtctcgagagacagcgccacccctgcgcgtgtgctgc 720
QY 4945 gtttccttccagaacatctcatctcaacgctgtctaagagagtcgcgcacatctccagagacatc 5004
Db 721 gtttccttccagaacatctcatctcaacgctgtctaagagagtcgcgcacatctccagagacatc 780
QY 5005 ctgtcccgccagcagagaggggtctctgtcccgagagacacttcaactgtgagcggtgtgta 5064
Db 781 ctgtcccgccagcagagaggggtctctgtcccgagagacacttcaactgtgagcggtgtgta 840
QY 5065 gggttgcctggaacagcagcgacttcaacatgagcgaggtctctacagagcggtgta 5124
Db 841 gggttgcctggaacagcagcgacttcaacatgagcgaggtctctacagagcggtgta 900
QY 5125 gaggtctacaagaacctcatctccatctctgtgaagcaccacggtgatacaagaagctgtgccc 5184
Db 901 gaggtctacaagaacctcatctccatctctgtgaagcaccacggtgatacaagaagctgtgccc 960
QY 5185 gcggttcagcagcaactgtcagagagccttcacacaaatcaatgacagagttccggtgtg 5244
Db 961 gcggttcagcagcaactgtcagagagccttcacacaaatcaatgacagagttccggtgtg 1020
QY 5245 gagcgctgttctgagagctatcttcgcgtgtggtctctacagcgccaccttcggtgacctg 5304
Db 1021 gagcgctgttctgagagctatcttcgcgtgtggtctctacagcgccaccttcggtgacctg 1080
QY 5305 gatgagcagaggttctgtgtacaagagcactgatacgaagctgtgcagagatctcaacac 5364
Db 1081 gatgagcagaggttctgtgtacaagagcactgatacgaagctgtgcagagatctcaacac 1140
QY 5365 cggtgtgagaggtctctacagcagagagattgtgcagcagctcgtctgagattacaaagac 5424
Db 1141 cggtgtgagaggtctctacagcagagagattgtgcagcagctcgtctgagattacaaagac 1200

QY 5425 tcttaccctgttgacaaagtccaaagcttgactcaacaaagccttaacatccagatccgcat 5484
Db 1201 tcttaccctgttgacaaagtccaaagcttgactcaacaaagccttaacatccagatccgcat 1260
QY 5485 gtgaaaccttactttatctccacagagctcaagggccgggttaactaatttaacccgacac 5544
Db 1261 gtgaaaccttactttatctccacagagctcaagggccgggttaactaatttaacccgacac 1320
QY 5545 tatggcttcgcacatctcctgttctgtcacgaccttcaacgcccagatgtggcgcacagg 5604
Db 1321 tatggcttcgcacatctcctgttctgtcacgaccttcaacgcccagatgtggcgcacagg 1380
QY 5605 gagctgcggagcaacacagcgttaagagctgtgtccagacagacacacagccttccctac 5664
Db 1381 gagctgcggagcaacacagcgttaagagctgtgtccagacagacacacagccttccctac 1440
QY 5665 atcaagactcgcataccgtgtgtgtccacccggagagacggtgtgacgccaagtggagtg 5724
Db 1441 atcaagactcgcataccgtgtgtgtccacccggagagacggtgtgacgccaagtggagtg 1500
QY 5725 gccatcgagagacatgcagaagaagacagcggtgcgttcctgtccaccgagcagaccca 5784
Db 1501 gccatcgagagacatgcagaagaagacacggtgcgttcctgtccaccgagcagaccca 1560
QY 5785 ccagatgtctaaagatgtacagatgtgtcttcagggtctgtatgggccacacgtgaaccag 5844
Db 1561 ccagatgtctaaagatgtacagatgtgtcttcagggtctgtatgggccacacgtgaaccag 1620
QY 5845 ggtccctgtgaggtgtgcccagggtgttttaagcagagatcccgaaagcccaagctcttc 5904
Db 1621 ggtccctgtgaggtgtgcccagggtgttttaagcagagatcccgaaagcccaagctcttc 1680
QY 5905 cggcatcacacaacaaatttggtgctctgcgttcaagggcttctgaagaaatgtgaagatgcg 5964
Db 1681 cggcatcacacaacaaatttggtgctctgcgttcaagggcttctgaagaaatgtgaagatgcg 1740
QY 5965 ctgtcggaanaataagggccctgtatgtggccggacccagaaagatcacacacgtgagctggag 6024
Db 1741 ctgtcggaanaataagggccctgtatgtggccggacccagaaagatcacacacgtgagctggag 1800
QY 6025 ccgaacttactgcgcctgtgagaggtctctgcagcccttgcttaaccacagcgtctgcccag 6084
Db 1801 ccgaacttactgcgcctgtgagaggtctctgcagcccttgcttaaccacagcgtctgcccag 1860
QY 6085 ctgattgcacccacccaccccgctctcagagaaactcttgaacagagcaagtttcggaag 6144
Db 1861 ctgattgcacccacccaccccgctctcagagaaactcttgaacagagcaagtttcggaag 1920
QY 6145 gcagaccttgcagcccaaggaacaaagctgttactctgaaggaacagcaccgggctc 6204
Db 1921 gcagaccttgcagcccaaggaacaaagctgttactctgaaggaacagcaccgggctc 1980
QY 6205 agctgtctgtgtctgagagggaggtctgcacctgtgtcccaactgtgtgtgggtgtgacaca 6264
Db 1981 agctgtctgtgtctgagagggaggtctgcacctgtgtgtcccaactgtgtgtgggtgtgacaca 2040
QY 6265 ctgtactgtgggtgtggccctctgcgcctgtgttcccaactctgtgtgcactatgttccct 6324
Db 2041 ctgtactgtgggtgtggccctctgcgcctgtgttcccaactctgtgtgcactatgttccct 2100
QY 6325 ccccttttaatttaaatgttttataagcaaaaaaa 6372
Db 2101 ccccttttaatttaaatgttttataagcaaaaaaa 2148

RESULT 3
AAH41934
ID AAH41934 standard; CDNA: 6828 BP.
XX
AC AAH41934;
XX
DT 30-AUG-2001 (first entry)

QY	1205	ccgagcagccttggcccaacatcgttgaagcagcgcctgggacgcttggaaacccggagctct-----	1253
Db	1226	caatccactttaaagaaatattgtttagacagtgcttgggaggttttggaaagaaattctacagaag	1285
QY	1258	-----gactcggagggcgagcgccgacccgtgagacagacccgccc-----	1300
Db	1286	tagaatactagttacttggagaaacgaaaagggtctcttggctcagaaggggaggaattcttagtatg	1345
QY	1301	---gtcgggggcccccaaggagccgggcgagtagttggggaaacgacgcttgcagcttcttcgct	1357
Db	1346	ttggcagaacgatacattctgaaagagacaacaagttggagatgatcttgttaacttgcagagct	1405
QY	1358	tcgcctccagcccaagcttaactgctacacaaactctcttaagaagaagcttgaagcactcaatg	1417
Db	1406	ttccagcacagctactcttcacagttgaacaattttttaaagcaggaagaagaaaccgtctaaatg	1465
QY	1418	acggagagaccccttcaaatgtctcttcgctcgagatactgagagcgccggctcgccctgcggcgac	1477
Db	1466	atgagaggtctcttacaatttcttcgttcgcgtgatactgagaaagcgacattcttcttctaaggcagc	1525
QY	1478	taagctcgttgaatgcgccagcttcaagaatcacatctcttcggctctcttgaataatcccaact	1537
Db	1526	taagaaacctattacagcttcacagctcaagatagacattcttcgcgacacttgaatatcccaact	1585
QY	1538	tcgtgcctctcccttggagctgctgtcttccatacaaaagccctacccggaaaccccaagggcccggcca	1597
Db	1586	atgctgccttcaactctccggagctgtgcttcaaggtbaaagcttctacacgttaagtagatltgaaccta	1645
QY	1598	ccaaaggagatcttcggaggtctcccgccggcgagagctcatgtacgtcccccatacccgctacagaa	1657
Db	1646	ccagagagaactcttagaggttcccgcaagggatggtttatgtttccaaacactactacagaa	1705
QY	1658	acctgtctgaagtgttaacccgacacagccctcaacttcaagcaagccggccagggctccgttgcgca	1717
Db	1706	atctctctctacataaccctccagagctcttaattttgccaatcgttcaagttctgtgaagaa	1765
QY	1718	acctgtgtgctggaggtgcagatgaatgacagggcgagagaccccgacagagctctgcgcgtca	1777
Db	1766	ataaataacaggaaagttccaggttttatgttatggagagagatccaaagcaatgcattgcgggttaa	1825
QY	1778	tcctttgcaagatccagctcagctgaatattacccggcgagagccttccacaggttggcttacc	1837
Db	1826	tcctttgtaaatcttagctgttccagaaatttccaagaagacattacagccgttagtatactc	1885
QY	1838	atacaaaagtcgcccgaggttctacagagagatttcaagctgatatcttccagcttgcgttgaacag	1897
Db	1886	ataacaggtctctcgatttctcaatgaaagaaatcaaggttlaaggtctctcttactttaactg	1945
QY	1898	agaaaccaatcaactgcgttctcaactcttaccatgtcaagcttgcacagcccgccggcgcaatg	1957
Db	1946	accatacactctgtccttcttacttcttcttaccatgtttagtgttcaaaaacaaaataactctc	2005
QY	1958	cccttggagaaacccgggtgccttacttggatcccaactcttgaagacagggcgaccttggaga	2017
Db	2006	ctcttgaacaacacagttctgatatatactgataaccatgcttcttcagaatgagacggttgaaga	2065
QY	2018	ccggaccccttcttcccccaggtctgttgaacacagccggccggccacgctatctccgttcca	2077
Db	2066	cttggccaggttttcttgcagctctcaatttggaaaacacacacagcttattcttcttactgt	2125
QY	2078	caaccgatatgttgcgcttccgggacatgcgcttgggttgaacggttccacaaggcggttctcagtg	2137
Db	2126	ctctcttgaagttctctctacactctgcaatgaatgttgtagataatcacaaaggtgttttaaty	2185
QY	2138	ttgagatctcaacagcgtgtctctcttggcaacccccagaaacccctactcttgaacaaattcttca	2197
Db	2186	ttgaagattgttggtttcttgcctcatccatacacaagaagatccctatcttgcacaatttcttg	2245
QY	2198	cccttgttgcacggtccgggagggagagccttcccaattccggctccaagaaacacttgtgtga	2257
Db	2246	ctctgttcaatgtctcgtgattgaacacactgttctccaggtcccaatttggggacatctgcgaatca	2305
QY	2258	ggcgaaggccaacgttgaagcagagagcttgcgggcagatcttgcagacatgagccttggccagcc	2317

Db	2306	tggaataataacttagaanaatgaaatgaaagagagatattcaagcaactgaaatcatccagc	2365
QY	2318	ccgaacccctcttgagctctcccccacagcgtctggaacagctctgtgcgtgcacga	2377
Db	2366	tggaaacagatgttcogattctctctcatctctgtctgataaactgatacttttagtatta	2425
QY	2378	ggccccgcatacagtgagcagatgtgtgaaccctgggccgtggagccttggaaagcaatg	2437
Db	2426	gaactccctgcatgtgtggccaataagtaagtaacctagtgcaagcactcttgaagccatg	2485
QY	2438	cccatgttaatcagccttgttcaacggagcgtggaagccagccaggaatgcocgcgtact	2497
Db	2486	catcaatataaatcgaactccaanaaacttggaaaggaatcatgacagacatgycagaa	2545
QY	2498	gcccaagatctgtctgcctacgctccacatacgccttccgcctctctgacacatgagccagcc	2557
Db	2546	acagccctcttgatcatatataatcatatgatttccgcgtaccacaataacttaacctaat	2605
QY	2558	tcocgga-----tggggccctccagtgacagtgcaagcgtgcacacatgycgcgtg	2608
Db	2606	catcatcacagatccctgggggtttggagagatgcatatgatgccaatgscatgtagat	2665
QY	2609	gctctgtgcgcccgagaagccctctacccgtgagcggttccaagagcatatcagcagcaacc	2668
Db	2666	ctgcgtgtgagaccttcgaagccttaatttcaatcgtcttcgaagcctttagtaatacaatc	2725
QY	2669	ctgacctgcgcgtggccccctggtcctctgtgagtagcagagt-----	2709
Db	2726	cagataatactgtgactcccacgataccagatgagaagtgtgatacatatcgaggata	2785
QY	2710	-----	2709
Db	2786	agggtttagatcgtcccaattccctgtgttaacactgtgttccaaaagctgcccatg	2845
QY	2710	-----	2709
Db	2846	gatccaaaccccgatcccaagtgacagaatacaacagagctatgtatcgaagtgtgaatcgt	2905
QY	2710	-----tcocgacatccgtggca	2725
Db	2906	tgtcttcgacacagagaagatcaagtcttcttcaacaacttaacggaagcgttacaacta	2965
QY	2726	gcaagctgtctcaagagagagctgtgctctgtcagtggtgtgtcgaacagacgtgcgtacg	2785
Db	2966	aaaagcttttcaacagagagcgtgctcttgacagtgggtgtgtgtcagtggtgcagcgttcgg	3025
QY	2786	aggccaactccagacagcgtgtctctctctccacagctaatgtgtgaagatgatgagcgtgc	2845
Db	3026	aatcagctttgcaacaagagcgtgtctcttcttgaataatggttaaaagagcatgtgtcaccc	3085
QY	2846	acctctgtcttggccagcagactagacacacaccccgcaagcgtgcgtctccccggagcgtcc	2905
Db	3086	attataactttaatgataaaacttggagctcccaaggaanaagtcgttttccagaagcattca	3145
QY	2906	tggacagacatacctgctctgtgtgtgtgtcgtgtgtggagtgagtgatcaaccggtgtccca	2965
Db	3146	tgtatgacaattgacagctctctgtgcagcaagatgtcagtgatatagtttcaacgatttcca	3205
QY	2966	aggatgttgagctgtgcccgaagcaactcaaacgcccagcctgtgcttcttccacgtgaacctc	3025
Db	3206	aggacaacagaatgtgtgagagacatcaatacaagcctgtgactcttcttcaatgatctgt	3265
QY	3026	tgtccctgttgagaccggggcttgtcttccagcgtgtgtccggggccacatacaagaagtggtg	3085
Db	3266	tgtctgtataggaacagagagatttgtttttagcctataaaatccgcgtcgtataaagaagtg	3325
QY	3086	ccagcgcgtccagatcgctccccaataccagaagccctgtgacctgtgcatagtgaatcca	3145
Db	3326	cttcaaaagcttaccatcattaccgaatcccagatgtctgtgttcccttgaagctgtgatttcc	3385
QY	3146	ccggaatccgtgtgacgacagacatctgtagcccttaaacctccctgtgtgcacctgt	3205

Db 3386 tacgaatcattctgcagtgatgagcactatggtatacataacttaaccttaacctcagcttactcta 3445
QY 3206 caactcagccttgcctccctctgtbctccacacacccctccagagctccactctc 3265
Db 3446 ctccactcattccatccactctctctctcctcgaacatactcagagcttcgtgtttt 3505
QY 3266 ccagccaagcccgagccccaaggtgacacagctgtttcgaaactgagtgacattccacgc 3325
Db 3506 ctacgaattgtaacagacaaagatgcaaatatgtttcaatatccgtcccttcgcgc 3565
QY 3326 agcagacttccctagctggtctcctgacagagctgcaactgcccgcgaacccgag 3385
Db 3566 aacagatattattggcagagactgtgttaacagagctggtctgatttaagaccctgag 3625
QY 3386 ctgaaggggacattcctgttgcacaaagggcaatcagtgctgtgcacagctgtatgtg 3445
Db 3626 ctgaaggagactgtttgattgataaagaagttcataatatgtatacaaatcttactctca 3685
QY 3446 gcatgaacattgaccccgctacgacgagggccactgtgaaaggtctgtgtgacagctgt 3505
Db 3686 gtacagactacagaccgaggtactcgtcacccctcaagaataagagctcgagtgccattgt 3745
QY 3506 acctgcacactgtcttcgatgtcagggatacccttgccacggtcgcatagtactgtcgtag 3565
Db 3746 atctactcctgattgttatattatcatgaaactgttacctgaagctgtatgtatttaacgaa 3805
QY 3566 gccacagtgacggtcgtacaagaagctcctcaatgctgtactcagaacagagagcggaagg 3625
Db 3806 ctacacatacacacaggaagac-----caattgtataccactgtatgattatgaagtgt 3859
QY 3626 acattgcggtgtacacaaacccctctgtgccaatgacattgtgtgccccctgaagcc 3685
Db 3860 agagcggaagtatgataaagccagacccgtgtgcacatgcgaagagacatcgtgtccccc 3919
QY 3686 ctggtctccgggcccagacattcccaggggcccacacaggtcttcgcgcaggtcgtgcc 3745
Db 3920 aactaacaaagcctgagcaattctcctccacgftcaacagagtgccaggaacacataact 3979
QY 3746 tctctctgtagtataagccggaacttgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3805
Db 3980 ttccacagagaataacgtcgaagccttcttcatcgttactgtgttctccaaaatgacg 4039
QY 3806 agccggtcgtcctcgtcagcggtgtgcccactgtacactccccccagcgtggagctgtg 3865
Db 4040 atgaacaagttcttcaacagaagtgttttaacagattctcagttcttcgcagctaaacgctat 4099
QY 3866 tggactgtgttaccttgcctgtgcgtctgtgagtacaaaggggaaagagcccttgaac 3925
Db 4100 taagatctgttcttctctgt 4159
QY 3926 gcatcaacagcctcacattccaanaaatctctgatatgaagggcggtcagaggaagcca 3985
Db 4160 gaattgaattagcttgacctttaaagaatacaaaagatcagagaagagcttgaagaagtca 4219
QY 3986 ttctgtgtacatcggagctcgcagaaagatgtgttcggagagctgtg----- 4033
Db 4220 ttcttgggagctaggtgtccagggcaagaatgttgaagcggaagccgaggaacagcttcgaa 4279
QY 4034 -----agaggaagccggttggagatccagagaacgtgcgtgtgcggaagagcttca 4084
Db 4280 gaagcccatctggaagtgtccttvgaaagtaagaataatttggaggtgaggaagaagtatga 4339
QY 4085 cacactggagaaactaactaagaccggtgagacaaagccaagatgaatvggaacacagag 4144
Db 4340 ctcaactggtctcaaaaacagagaagctgtcaaaatcaagaacagagatttgaacacagag 4399
QY 4145 ccttgtgtgaaggaacctgtgcacacagaggaagccctgaatgtgttcgcgacacatgaga 4204
Db 4400 caactgtgtatggaacacgtgcgtacacagaagcaacctaatcatatttaagtaactatagaa 4459
QY 4205 tcatctgtcagagacgtgtgagtcttcagaagcccggaagagcgtcttggggcagtgctga 4264
Db 4460 ttgttgtacagaccgtttctgtaaacggaatccaagaagagcattctgtgtgagtgctaa 4519

QY 4265 aggtgtgtctgtacagcctgggcagtgcccaagtgccctctctctgtgaagatggcctg 4324
Db 4520 aagtgtaactacacagcagtctgtttaaaccaagtgcagtttactatcaaacactgtttg 4579
QY 4325 ccaccaagagggcccttgtgttccaagttcccgagctgtctgttctgagaggaagacagagc 4384
Db 4580 ctacacagagagcccttgttccaagttcttcgaactcttatttgaagaagagacagagc 4639
QY 4385 tgtgtgcagacctgtgtcgtgaggtccctacagacactgtgagcgccgaatccagccatcc 4444
Db 4640 agtvgcctgatttaagctctcaggtcttcgcgaactgtgagcagatcagatcogtgcataac 4699
QY 4445 gcaacgcgcagcgccctcgtctgtatcctgtcattgtgcagacagaacttcgagatcggccaca 4504
Db 4700 gtccacacccagctgcctccctcttaccataatgagcaaaacttvgagttvggaata 4759
QY 4505 actttggccggtgtgaagtgaggtgcacatgtctctctgtccctgtgtgggagcagcgc 4554
Db 4760 actttgcaggggttaaagatgaggttaaccaatgtacattccctctgtgtgggcacatctc 4819
QY 4565 agaactcagtgaaagagacactgtgcagcgttccactcaaaaacatccctccactatgtag 4624
Db 4820 agaatttaagaaagatctctaagacgtctctctaagacataattgaacatgtctgaag 4879
QY 4625 aggacatvgggctgcgggacagcccttcgcagagcaggttcacaggaacctgtataacc 4684
Db 4880 aagactgtgaattgaggaagaaacacattctcctgacaggtccagagatcgtgtttacatc 4939
QY 4685 tgcacatgatccctgcagacacaggtgaaagatgaaagaaacacaggaagacccctgagtgc 4744
Db 4940 tccataatgtctctctctgtatctgtgaanaatgaaagaaacacaggaagaaatccctgaatgt 4999
QY 4745 tcatcgacctatgtacagaaattgcaccgggtgtaccaggggtctcacccgacctgcgtga 4804
Db 5000 tgattgtactaatgtacagaaattggcaaggtttacagacactctccaga---gcatgtga 5056
QY 4805 cctgtgtcagaacatctgcccgggaagcgcgagagctcgggaacacacgcgcgagccgc 4864
Db 5057 cctgtgtcagaacatctgcccgggaagcgcgcacgtcgaagaacaaatcatcgtgaactgtcac 5116
QY 4865 agtgcattgtcagcgggcgcgcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 4924
Db 5117 agtgcattgtcagcgggcgcgcactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5176
QY 4925 gccacctgcggtgtggtgcgtcttccctccagaaacatccatccacagctgtcagaagagt 4984
Db 5177 aatatcttcctgt 5236
QY 4985 ccgcacatctccgagacatccctgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 5044
Db 5237 ctgcggtctccagatgattgtgttatctccagatgaagaagatctgtctcgtgaataact 5296
QY 5045 tcaactgagctgtggctgt 5104
Db 5297 ttacagatcagagactgt 5356
QY 5105 ggtcttcagagcggtgtgaatgaggtcttcaagaacatccatcccatccctcgtgaagccacc 5164
Db 5357 gcatgtatgaagcagtttaagtttaagaatttacaagatctattcttcatcattgaagctaatc 5416
QY 5165 gtgcataaagaagctgtgcgcgcgtgtgcacagcgaactgcgcgcgcgcgcgcgcgcgcgcgcgc 5224
Db 5417 gggatgcaaaagaacttcccaattcatctgtttaaactctcaagaagacatltagaanaattg 5476
QY 5225 tgcacaaagattcccgctcggagagcggtgttcggagacgtatttccgcgtgtgtgtgtgtgt 5284
Db 5477 ttcatcagagtaactgcgtcggagcggt 5356
QY 5285 ggcgccactcgtgtgaacctgtgaatgagcagaagattgtgtataaggaagacatcgatacga 5344
Db 5537 gaaccaagttcgggatttggatgtacaagaatgtgttacaagagccgtgcgaataacga 5596

Db 2109 ccaaggacaacacactggaagatcttcaacccctctgcacccctctggagagcgagcgac 2168
OY 2226 ctcccatctcgcgctcaaggacactgctgagaggaagcgcaacgttgagagcgagcgcg 2285
Db 2169 ctcccatctcgcgctgctgagatcaagaataatcagacgaatgctgcgttgagagatgagcgaa 2228
OY 2286 ggcagctcttgcaagcactgacgctctgagccagcccggaacccctctgtgacctcttcaccca 2345
Db 2229 gctcagacatcatctgcggaactctcccgctcgtagcgctcgctctctccgcacct 2288
OY 2346 cgtctctgacaagactctgtgctctgctgcatcaaggcccccgaatcaatgagtgccagatgt 2405
Db 2289 ggtctctgacaagactctctccagcgtctcgttgacagcccaatgcatcgcgtcgagcagacag 2348
OY 2406 gaacctggccgctgagagcctcttgaaagcaatgagcccaatgtagaagcttgtaaccggag 2465
Db 2349 caactctccacgaattctgcctctgagatccggtggtgagccatcgcaacagcttgacaacag 2408
OY 2466 ccttgagagcagcccaagatgcccgcgttcaactgccaagctgctgactacatca 2525
Db 2409 caagacctgagcagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2468
OY 2526 cgctcttcgacctcttgcaactgagcccaagcctcccgatgagggccctccagtgacagt 2585
Db 2469 cgctctccgcctcgccagaggtgacaagggatgtgcccgaatcgagcgtctccactgacct 2528
OY 2586 gcaagctgcg-----caacatgcccgcgtgctctgctgctcgcgcccgcagccctcta 2633
Db 2529 ccttagacctcgcgagctacacacacgtatggtcgacatccagctgcgtctgagtgatcca 2588
OY 2634 ccttgccgcttccaagaagcatcagcagcagcaacccctgacccctgcgcgtgagccctgagctc 2693
Db 2589 gctgtctgacaagcccggttgatgagcagcagtaaacccagacctcgcgtgagcagacatccgc 2648
OY 2694 tgtgatacgaaggcttcccgcatccctgagccagcagctgctccaagagga----- 2744
Db 2649 agcagagcagagagtgagaaacatcatgcttcaaaagatccgcagatccgcagatcgcaagccg 2708
OY 2745 -----gctgagctctgcagtgagggtggtcagcagagctgagccgtatgcgtgagccg 2790
Db 2709 aatgtactatattgctctcgtagtgatgctccaagttaaccccgagcccaagcgcc 2768
OY 2791 atccctcagcagcgct-----ggtctcttcacagctcat 2825
Db 2769 agccagagaaacatctccaatgagagcttgcccttcagatgtgttcagcccggaat 2828
OY 2826 ggtgaagagatagtcgctgagcactgctgctgagccagcagctagacacaccccgcaagct 2885
Db 2829 ggtgaagaaagatggtccagcagcatataacatgagcaaacggagacagtttcgagagac 2888
OY 2886 ggcctctcccgagagcttctctgagcagacatcaactgctgtggtgctctgtgagccctgga 2945
Db 2889 tctgttcttcgagccgttcatatgacataactatattgttaatgtgttcacctcgga 2948
OY 2946 ggtcatcacccggt-----ccaagaagatgtgagcgtgagccgagcaacactcaagc 2996
Db 2949 aattgcagccctttagtaaaaccacagaagaataatgaacgcgggaataagatgacat 3008
OY 2997 cagcctgagcttctccatgtagccttctgctcgtgagccggagcgtctgttcttcag 3056
Db 3009 cagcctgagcttctctcttgatgacctctctcccatagatcgagggttctgttttaa 3068
OY 3057 cctggtccgggcccactacaagaggtgagcagcgctccagctgctccctaataccag 3116
Db 3069 cctcatcaagaacatattgacagcagctgtcagccaagct-----cagtaacctcc 3119
OY 3117 agccctgcgagaccctgcatggaatccaacccgcatccctgtgacgccaagacatagct 3176
Db 3120 aagcgcatcttccatgagcctagatctctgagaaatccctctgtaagcagacatcatct 3179
OY 3177 gacctcaacactccctgctgcccctgtcacctccagctcgacctccctctctgtctc 3236

Db 3180 caactgaaact-----tcttttatgaaatgctatgactgtctcaacatctctctgtcc 3233
OY 3237 ctccaccacctcccaagactccaactctctccagccaagcccgagcccaagtgacag 3296
Db 3234 ttcacatctccccaagaataatcaagctctccgcagctctccagacacagaaatcgccag 3293
OY 3297 catgtcgaaactgagtgagcccatcccggaagacatctccagctctgagctctctgac 3356
Db 3294 catgtctgactctgactctccgaatlaacgcgcagacagacatctccacccgggtctctccac 3353
OY 3357 ggaactgagcctgagccctcgaaacccctgaagcgttgaaaggcgcatctctgttgcacaagagcg 3416
Db 3354 agaactgctgctccctgatatgcagaaagggaaggaatcagcaagtaacaaaggaaagcg 3413
OY 3417 catcagtgcttgacagagcctgtatgtggtccatgaacatgaccccgctacgcggagcg 3476
Db 3414 tgtcagtgcaatctcaacagctgcttaagtctcaacgacctggaacccagcgtgtgtcaaac 3473
OY 3477 cactggaagcctcgtgtgagccgagctgaacttgaccctgacactgctctgatatgacggagac 3536
Db 3474 agagtggaagggtcaaaatccgcgccttactactacttaagttagtcatcatcttgatgc 3533
OY 3537 cttgcacagcgtgcatagtactgtctgagggcccaaggtcagcgttcaagactgacctcaat 3596
Db 3534 ttgccaacagctctgtgacttatacagtgagaaatctcgagatactcgagataccgac----- 3584
OY 3597 gcttgatctacgacacagaagcggaaaggagacattgacgtgtccatccatccacctctgtgagc 3656
Db 3585 -----cagtgctcaggaagaaagaaagagagccggtgcacatcaacagaaatgtgagc 3658
OY 3657 catgccaatgtctgtgtgagcccccctagccccctgagcccgggccagacatcccgaggcg 3716
Db 3639 tctgccaatagaggaatlaattcaatttgaaacaagtgagatagt-----gctgctc 3692
OY 3717 acccaagcctctcgcgcagagcgtgtgcctctctgctgtatgaacggcgacctgtgcgcg 3776
Db 3693 ttccttgccctataagcagatcaacatgctgaacccggaacatctacgtccaacatcatgat 3752
OY 3777 gttgtgctgtgtgtgtgagaaacacccagccggcgtctctcctgcagcgtctgagccactga 3836
Db 3753 cgtctctctctgagatcaagaaatgctgatacgaagcctcaattgaaagtgtatgtctga 3812
OY 3837 cctgaacactccccagcgtggaagcgtgtgtgacttgctgaactttgctccgtgcgctt 3896
Db 3813 cctgcacatcaacgcagctcaacaggaatttagactacttctcatctgtgtgtattgtt 3872
OY 3897 tgaatcaaggggaaagagccttgaacgcataaacgcctccatctcaaaaatctctc 3956
Db 3873 tgaatcaaggggaaacagagttctcgacaaggtcagtaacccaagtctctgcaggaagtcag 3932
OY 3957 ggaatgaagcgcggtgagtagaggaagccatctctgtgtaacatctcgagactcgacaagaat 4016
Db 3933 ggaatgaagcgcggtgagtagaggaagcgttctgtcgttgaggaaaggccagaggagat 3992
OY 4017 ggttcggagagtcgttgagagagagccggtt-----ggaaatccagaagacgttgagctg 4070
Db 3993 gatgycgcgcgggtctccaagggaaacgcagatcttccaggtccataaagtgaatltgagatg 4052
OY 4071 gcggaagagcgtlccacaacttgaaagcaaaccttcagcagcggttggaacaggaagtgga 4130
Db 4053 gaagaaagagcagacacatgtgcggcaagctaatgaaagactaataaacaagagccga 4112
OY 4131 aatgaaacagagccttggtgagagggaaccttgagcaacccagagcaagctlagtgtctc 4190
Db 4113 gtagatcaagaagccttgatcagtgacatctgtgcatacgaagcacaatttaatacctc 4172
OY 4191 ggaacactgagatcatcgtgcagagcgtgagatcttcaagaaagccggagagcgtctc 4250
Db 4173 ggaatgcaagaaacaaatlaacagcagagcgtcgtctgtgactgtaagaacagcgtctc 4232
OY 4251 ggggagctgtgagagttgtgctgtaacgctgtggagggccagagtgccctctctc 4310
Db 4233 gggaggtgtctcagagggtgcgtgtaattctctgaactgtgatacagaggtacacactac 4292

QY	4311	gagcgatagcgctggccacccagagggcccttgctgtccaaagtcccgagcgctgcgttccga	4370
Db	4293	gactcaatcgttttgcaaacactccgtgtctctcatccgcgaagtttgtagacttactcttcga	4352
QY	4371	ggaagacacaggaactctgtgacccgaactctgctctgagagctccctacagacactgtgcagacg	4430
Db	4353	agagagaggtggaacagtggttccgaactatgtcaaccaagctccgcacacactgtgagcagag	4412
QY	4431	catcaagcaccaatctcgcaacgcagccagcgctccgtctgaactgtctcatctgcagagaact	4490
Db	4413	catgtagtcatcccgagaccaaagcctgttcacacactcttaccctccctacagaaagtctcagtt	4472
QY	4491	cgaagatcgccacacactgttcccgctgtgaagaatgcagagctcaacatgctctccctgcctc	4550
Db	4473	tggagccacacagtaatttttgcagaagtagtaagatgcagaagtaacatgctccctgcagctctt	4532
QY	4551	ggttgaggcagacgcagaaactcgaagtgaagaagacactgcgcgttcactcaaaacactctc	4610
Db	4533	ggttggaagagacacacagactcttaataagaagacactgcagaaagatcccttgagacaattt	4592
QY	4611	caactatgcctgaagagagacacttggtgcgtgggaacagacacttcgcagagagcgccagga	4670
Db	4593	ggcctatctcagaagaggaacccagccatgcagatgatctcttccaccocccaggtgagagga	4652
QY	4671	ccgtgatctcaaccttcacatagtatccctgcagacacggtgaaagatgaagacacacagga	4730
Db	4653	aactctctgaactctgaatgatgcactcttatatgacaacagtgaaagtaggagatttcagga	4712
QY	4731	ggagaccctggaatctctcatcgactctatgtacagaatctgcgcgggtctccaggtctacc	4790
Db	4713	agatccctggaatctcttatgtatctcatgtctatgcagaatctgcagaagatctccacagctctcc	4772
QY	4791	ggaacctccgctcagctcgtgtgtgcagaacaatgtgcgcaggaagcacgcggtgagctgcgaaca	4850
Db	4773	tgatctgcgctgactgcctggtctccagagaacatgtgcagagaacaacaccagaagaagtgtga	4832
QY	4851	cgcccgagcgcccgccagtgcatgtgtgcacgcgcgcgcctctgtgctgaagtactccgcct	4910
Db	4833	caacgaaagctgcacatgacgtgcgtgtgcacgcgcctcgtatgaagtgctgaatctctgaagat	4892
QY	4911	gctcgagaggaacacgcgcacactgcgcgtgtgcgtgttctctctccagaacatctcaatcaaa	4970
Db	4893	gctcgagagacacacagactcaactccgcgtgtgcaggtctcaagcttccagaatattcttccaaa	4952
QY	4971	cggtcagagagactccgcacatctccgcagacatcctgtctgcgcgcgcgcgcgcgcgccttg	5030
Db	4953	tgtgtctgagagactctgtgtgtgtctgtagagacacccctgtcacactgcagcgagatvggtgtg	5012
QY	5031	ctcccggaagacacttcaactgtagctgtgggtcgtgtaggtgtgtcgtgcgaacgcgcgcgcgcta	5090
Db	5013	cgacgcgcacatctcaaccgcagagatggtcgtgttagagctcctcgcgcgcgcgcgcgcgcgagact	5072
QY	5091	cttcaacatgttggcggtctctacagagtcggttgataggtgtctcaagaacatctcatcccat	5150
Db	5073	cttcaagcagcggcggtctctatatgagacagttaatgtaggtgtctcaagctcgtgtcatcccat	5132
QY	5151	ccgtggaagccacacgtgtactacaagaagctgtgcgcgcgcgtgtgcacgcgcacactgcagagagc	5210
Db	5133	cctagaagcgcatctcgaaatctccgcgaatctgcacatctcaacagcagctgcgcagagcgc	5192
QY	5211	cttcaacagaatctatgcacccaaggttccgcgtcgggagcgcggtgtctgcgaacgtatcttcc	5270
Db	5193	cttccgaagacatctgttaacaaag-----gatcataagaaagttgttggaactactctccg	5246
QY	5271	cggtgagctcttaccgcgccacactcgtgtgacctbgtatgtagcgcagaggtcttggtgtacaaga	5330
Db	5247	agttcgttctcttggatccaattgttgagattgtgtagtgcagagaggttgcgtctacaaga	5306
QY	5331	gcacatgatcaacgaagctgtgcagagaatctcaacacgcgcgtgcgagagaggtctctacaagagag	5390
Db	5307	gcttcgcacattacacaggtcttccttgaagcttcccaataaactacagagagattctatgtctaatg	5366

QY	5391	atttgcgcagcagcagctcgctgttgatattatcaagaactcttaccctgttgacaaagtccaagct	5450
Db	5367	ttttgttgcacaatttctgtgaaigtatattaagaactccactctgttgacaaaaccaagtt	5426
QY	5451	tgaccacaaaagcgcctacatccacagatcagctatgtgaaacggtactttgtactacga	5510
Db	5427	ggatcttcaacaaagcctcatatcacagatcaactttgttgagccctactttgttgatatga	5486
QY	5511	gtccaagaacccgggtgacactacttgaaccggaactatgtggtccttcgcacactcgttctg	5570
Db	5487	gatgaagaagcaggggtcacatactttgagaagaatttcaactccggaggtttcatgtacac	5546
QY	5571	caacgccttcaaccccggtatggcgcgcacacggggaggtgcctccgagcaaccacagcgtaa	5630
Db	5547	caacccgttcaaccccttggaggcgctccctcgggaaagtgtgatbgaacgtatcaaaagaa	5606
QY	5631	gacgtgtctcaacacacgacacgccttccctccatcaagaactcgatccgtgtgtgca	5690
Db	5607	cacagctctgacacatacgaacgccttccctcatcaagaaccggatcagcgtcatcca	5666
QY	5691	ccggagagagagacggtgtcgtcagcgcagctggaggtgccaatcgagagacatgccaagaagac	5750
Db	5667	gaagagaggaattgtttcttgacacacgattgaaigtccatgaaagacatgaaagaagac	5726
QY	5751	acggagagctgtgccttgcaccacggagcagagaccacaaatgtctaagatgtctaacgattgt	5810
Db	5727	cctgtcagtttagcagttgtgccatbaaccaggaagccgctgtatgaagaatgcttcagattgt	5786
QY	5811	gcttcaaggctcttgtagggccacacgcttbaacagaggtcccttgaggtggtgcaccagttt	5870
Db	5787	gtctcaaggtctctggtggaggtactacgttaatacgaagacacttggaaagtgaagcccaagttt	5846
QY	5871	tttaacagagatccccggagaagaccccaagctcttccggagcatcacaaaatitgcgctctg	5930
Db	5847	ttgtgtcgaatctctgcctgtcatcaaaaactctatcgatcatcaacaagattgagttatg	5906
QY	5931	cttcaagaagactctgcagaagaattgtgagatgctgtcgcggaataaagccctgatgtg	5990
Db	5907	ctttagaattcatcatcagatcagatgtgtggtgagactgtlaagaaaaaacacgcttcatcac	5966
QY	5991	ggcggacacgaagaggtaccacacgtgagcttgagcgacactctgcgcgtgcggagagc	6050
Db	5967	ggcagacacgaagggaatcatcagcaggaactccaanaagaactataaacaagctaaagaagaa	6026
QY	6051	tctgcagcccccgtcttaccacgcgccttgcacagctg	6087
Db	6027	cctcagcgccaatgatcgcgagcggaattccagaactg	6063
RESULT 5			
AAH99572			
ID	AAH99572 standard; cDNA: 1652 BP.		
XX	AAH99572;		
DT	16-OCT-2001 (first entry)		
XX	Human protein encoding cDNA sequence SEQ ID NO:407.		
DE			
XX	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;		
KM	antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;		
KM	antibacterial; endocrine; cardiant; central nervous system; virucide;		
KM	anti-HIV; fungicide; antimitogen; cardiovascular; antianemic; anemola;		
KM	antiaagregant; haemostatic; vulnerray; antulcer; osteopathic; eczema;		
KM	dermatological; antiallergic; antiaschmatic; antididiabetic; cytostatic;		
KM	neuroprotective; antidiressant; nootropic; antiparkinsonian; infection		
KM	immunosimulant; gene therapy; antisense therapy; vaccine; inflammation		
KM	antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;		
KM	cardiac dysfunction; neuropathologic disorder; cardiac anaphylaxis; autoimmunity;		
KM	genetic disease; haematopoeitic disorder; platelet disorder; asthma;		
KM	thrombocytopoenia; osteoporosis; severe combined immunodeficiency;		
KM	allergic rhinitis; diabetes; multiple sclerosis; depression;		
KM	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;		

KW neurological disorder; ss.
XX
OS Homo sapiens.
XX
PN M0200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000MO-US55017.
XX
PR 23-DEC-1999; 99US-0471275.
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
XX
PA (HYSEQ-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT.
XX
DR WPI; 2001-457603/49.
DR P-PSDB; AAM25631.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
PS Claim 1; Page 505; 1217pp; English.
XX
XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antihaemic; antiaggregant; haemostatic; vulnereary;
CC antihyper; osteopathic; dermatological; antiallergic; antiashtmatic;
CC antidiabetic; cyostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders.
XX
XX
SQ Sequence 1652 BP; 377 A; 478 C; 479 G; 318 T; 0 other;

Query Match 23.0%; Score 1465.2; DB 22; Length 1652;
Best Local Similarity 99.7%; Pred. No. 9.3e-287;
Matches 1489; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 4879 gggcgccgctcggtgctgagtaactcgcgcctgctcgagaccagcgccacctgcccgtg 4938
DB 2 gggcgccgctcggtgctgagtaactcgcgcctgctcgagaccagcgccacctgcccgtg 61
QY 4939 ggttggttcccttcagaaacatcatcaaaagtgtctagagagtcgcgcacatctccgac 4998
DB 62 ggttggttcccttcagaaacatcatcaaaagtgtctagagagtcgcgcacatctccgac 121
QY 4999 gaacatctgctcgccagcagagagaggtctctgctccgggaaacaacttaactgagtcgg 5058
DB 122 gaacatctgctcgccagcagagagaggtctctgctccgggaaacaacttaactgagtcgg 181
QY 5059 ctggttaggggttgctgtaacagcagcggtacttaaccatggtgcgggtcttaacagagcg 5118
DB 182 ctggttaggggttgctgtaacagcagcggtacttaaccatggtgcgggtcttaacagagcg 241
QY 5119 gtgaatgaggtctacaagaactcatcccatccttggaagcccaacgctgactacaagaag 5178
|||||

DB 242 gtgaatgaggtctacaagaactcatcccatccttggaagcccaacgctgactacaagaag 301
QY 5179 ctggcccggtgtgcaagcagaacatgtcagagagccttcaccaagatcatgcaacaagtctc 5228
DB 302 ctggcccggtgtgcaagcagaacatgtcagagagccttcaccaagatcatgcaacaagtctc 361
QY 5239 ggttggaagcgtgtgttcggagcgtatttcgcgcgtggtgcttcacgggcccactcgtgt 5298
DB 362 ggttggaagcgtgtgttcggagcgtatttcgcgcgtggtgcttcacgggcccactcgtgt 421
QY 5299 gacctgtgtagcaagagttgtgtlacaagagccatcgaatcaagaagctgtgcaagatc 5358
DB 422 gacctgtgtagcaagagttgtgtlacaagagccatcgaatcaagaagctgtgcaagatc 481
QY 5359 tcacacccggtgtggaagatttctacacgagagatttggcgcgcgcgttcgttgaaattatc 5418
DB 482 tcacacccggtgtggaagatttctacacgagagatttggcgcgcgcgttcgttgaaattatc 541
QY 5419 aaagacttcaacccctgtgacaagaagtcgaagcttgactacacaagaagcctacatccagatc 5478
DB 542 aaagacttcaacccctgtgacaagaagtcgaagcttgactacacaagaagcctacatccagatc 601
QY 5479 acgtatgtggaacgcgtacttgatacctacagactcaagaagcgggtgactacttgac 5538
DB 602 acgtatgtggaacgcgtacttgatacctacagactcaagaagcgggtgactacttgac 661
QY 5539 cgcacacttggtcgttcgcacatctcctgttctgcacgcgcgttcacagcggatgggcgcga 5598
DB 662 cgcacacttggtcgttcgcacatctcctgttctgcacgcgcgttcacagcggatgggcgcga 721
QY 5599 cagcggagctgtcccggaacacacacaaagcgtlaagacgcgtctcaagacacgaacgccttc 5658
DB 722 cagcggagctgtcccggaacacacacaaagcgtlaagacgcgtctcaagacacgaacgccttc 781
QY 5659 ccctacatcaagactcgcgtatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5717
DB 782 ccctacatcaagactcgcgtatcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 841
QY 5718 ggaagtgcctatcgcagagcatgcagaagaagacacgagagctgtgcttcgacacagga 5777
DB 842 ggaagtgcctatcgcagagcatgcagaagaagacacgagagctgtgcttcgacacagga 901
QY 5778 ggaacccacagatgtctaagatgtctacagatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 5837
DB 902 ggaacccacagatgtctaagatgt 961
QY 5838 gaaacagaggtccctcggt 5897
DB 962 gaaacagaggtccctcggt 1021
QY 5898 gctcttcgcgcatcaacaacaattgctgctgtctgtcaagagactctgc-aagaatgtg 5956
DB 1022 gctcttcgcgcatcaacaacaattgctgctgtctgtcaagagactctgc-aagaatgtg 1081
QY 5957 aggatgtgcgtgcggaataaataagagccctgtatgtggccgagacagaagagatcacacctg 6016
DB 1082 aggatgtgcgtgcggaataaataagagccctgtatgtggccgagacagaagagatcacacctg 1141
QY 6017 agctgtgagcgaactactgcccgtcgtgcggaggtgtctgtcagcggccctgtcttaacagagcc 6076
DB 1142 agctgtgagcgaactactgcccgtcgtgcggaggtgtctgtcagcggccctgtcttaacagagcc 1201
QY 6077 tgcgccagatgtagcacccacccaccccgagcctcaagaaactccttgaagagcaagtt 6136
DB 1202 tgcgccagatgtagcacccacccaccccgagcctcaagaaactccttgaagagcaagtt 1261
QY 6137 tcggaagagcagacacttgtagcccaacaaggaaccaaaagctgtlaactlaagaggaacagacac 6196
DB 1262 tcggaagagcagacacttgtagcccaacaaggaaccaaaagctgtlaactlaagaggaacagacac 1321
QY 6197 cgggcctcaagctgtctgt 6256
DB 1322 cgggcctcaagctgtctgt 1381

OY	3673	ggccccatgacccctggtctccgggcccagatctccaaaggccaaacggctcttcgc	3732
Db	1219	acatctggctctcaactcaacaagaagccttgcgagtttccctctcaactcaacgattgcccag	1278
OY	3733	gcagggctggccctctctgtctgtatgaagccggagcctctgcggctgtgtgtgtgtgtgtgt	3792
Db	1279	caaacactaactctttcaagaagaatcaagtctgaagcctttgatctgtcttaactttggctt	1338
OY	3793	ctgaaaaaacccgagcccgcgctctctgcgcgtctggccacatgcatactccccag	3852
Db	1339	ctcaaaaatgcagttgaacacgttctacagaagtgtttacagatctctcagcttcttcag	1398
OY	3853	ctgggaagctctgtttgagctcttcgtctatcctttgcctgcgtgccttttgattgaaggggaa	3912
Db	1399	ctaaacccggtatatagtctgcgtcttactctctgtgtcttctgtttctgtgaatataaaggaa	1458
OY	3913	aagcctttgaacgcataccaacagcccaacttcaaaaatctctgatatgaaggccgg	3972
Db	1459	aagatgttgaacgaatgaaatgacttgcaccttaagaataatcaaaagacatgagtgcaag	1518
OY	3973	ctagaggaagccattcttgggtacacatccggagctgcgaagaatgtgttcggcaagtcgt	4032
Db	1519	cttgaagaagctattcttctgggagcataggttccagggcaagaatgtgtacggggaagccga	1578
OY	4033	g-----agaggagccgcttttgggaatccagagagaactgctgcgcg	4071
Db	1579	ggacagctcgagagaagcccatctgtaagtctccttctggaagtcaagaataatttggatgtg	1638
OY	4072	cggagaagcgctcacacactctgtaagcaaaccttcagagccggtgtgtaacaaagaccaagttaa	4131
Db	1639	aggaaagatatgactcaactctgcgtctcaaaaacacagaaagcttgcacaatacaagtgcaag	1698
OY	4132	atggaaacagagccctgtgtgtaagayggaactgtgcaaccggaggaacgctagtggtctg	4191
Db	1699	attaaacaggaagcactgtgattgtatggaacactgtgtacagaagaacaaacttaactttaa	1758
OY	4192	gacacacttgagatcatcactgcgcagaggtgtgattgtcttcagaagcccggaagagcgctctg	4251
Db	1759	gatacatatagagatctgtctcttcagaccggttctctgttaaaggaaatccaagaagagacttct	1818
OY	4252	ggggcagctgcgtgaaggtgtctgtcgtatcaagcctctgcagctgtgcagtgccactctctg	4311
Db	1819	gggtgagtgctaaaagagtgctacttacaacagatgagcctgttaaccaagaatgtgcagttacta	1878
OY	4312	cagcatgtgctgtgcacccagagggccctgtgttccaaatctcccgagctgtcgttctcag	4371
Db	1879	caaacactgtttgtctacaacagagagcctctgtgttccaagaattctcgtgaactcttatttaa	1938
OY	4372	gaggaaacggagctgtgttcgcgcagccgtgtcgtgtaggctctctacggaaactgtgtgcagccgc	4431
Db	1939	gaagaagacagagcagtggtgtgtattatgtccttaagctctctccgaacactgtgaagtgtagc	1998
OY	4432	atcagcacactccgaacgacgcgcagcgcctgcgtcttactgtctatgtcgaagaacttc	4491
Db	1999	atcgtaataatacgttcaacccccagtgctccctctactactaataatgaggaaaactt	2058
OY	4492	gagatcggccaacatttgcgcgtgtgtgaagtgcaggtcaacatgtctctctgcctcgt	4551
Db	2059	gagattgtgataaacttctgcaggtttaaaatgcagytacaatgttcaacttccactcttg	2118
OY	4552	gttggggcgagcgcgaactctcagtgaagagcacctgcgcagcttcaactcaaaacactctc	4611
Db	2119	gttgggcacatctccgaattttaaigtgaagaattcttaagcgttctcttcaaaagctatattg	2178
OY	4612	accatctctagagggacatgtggggctgcggggaagaacactctgcgaagcagatccaggac	4671
Db	2179	acatatgtctgaagaagatctcggaaatgtgaagggaacaacacttccctgataccggttccagat	2238
OY	4672	ctgattgtcaacctgcacatgatactctgcagcaggaacggttgaagtgaaggaaacaccagag	4731
Db	2239	ctggttttcaatctccatagattcttctctgatactgtgaaaatgaaggaaacaccagag	2298

QY	4732	gaccttgatgctccatccagcccatctgtaaagaattgccggggtctaccaggtctccagc	4731
Db	2299	gaccttgaaatctgttatctgtatctcaatgttaaagaatttgcagaaggtgttaccagacctcca	2358
QY	4792	gacctctgcagctctgtgttgcagaacatbgtccgggaagaacacgcggaggtctggcaacc	4851
Db	2359	gactctgcgattgaccttggttgcagaacatbgtccgggaacatctccgaatcgaagaagcaatat	2418
QY	4852	gcggagagccgcgcagctgcatctgtgcacgcggccctccctgtgctgaatccctgcgcct	4911
Db	2419	gctgaagctgtgcacagtgctctagctccactcagcagcaattgtgtcatatttggacatg	2478
QY	4912	ctcgagagccacggccgcacccctgcggccgtgggtgtgttctctccagaaactccatccac	4971
Db	2479	ctgagagagcccggaataatctctccgttggagctgtgtaacattccagatatatttccatcat	2538
QY	4972	gtgcctgaagaggtctccgcatctccgaagaatctgtctgcgccgaagagaggtctctgc	5033
Db	2539	gttttagaagaatctgcgtgcctcagatgatgtgtatctccagatgaagaagtatctgc	2598
QY	5032	tcggggaagccactcactctgagcttggtgtgtgtatgggtctgtcgaaagcgagccgctac	5091
Db	2599	tcgtgaaataacttacttactgtgcagcttgcagcttgggtatctactgaaacaagcagtctcc	2658
QY	5092	ttcaaccatgggcgggctccctacagagcggtgatagagctctacaagaactcaccctac	5153
Db	2659	ttctctcatagctctgcacgttatctgaagcagtttaatgaaagtcttacaagaattacttccat	2718
QY	5152	ctggaagagcccccgtgactctacaagaagctctgcgcggtgtcaacgcgaactctgaaagagc	5211
Db	2719	catgaagcttaatctgggtgtgcagaagaactctccacaattcatggttaaaccttcaagaagca	2778
QY	5212	ttcacccaagatcatgacacccagagttccgcgtctggagcgcggtgtctcggaagatattccgc	5271
Db	2779	ttcagcaaaatgtctcatccagagtaactgcgtgtggagcggatgttgcacattatttcgt	2838
QY	5272	gtggagctctctacgcgcgccactccgtgtgacctgtagatgaacagagatttgtgtacaagag	5331
Db	2839	gttggttttttatgtaacccaagattctcggtgatttggatgtgaacaagaatttgtttacaagag	2898
QY	5332	ccatcagatcagaagcctgcgacagagatctccacccggtctgaaagagtcttccacccgaagga	5391
Db	2899	cctgcataaacccaactctgcagagatatactcccaagatttggaggtatttaacggtgaagaa	2955
QY	5392	tttggcgagcagctcgttttgatgathatccaaagactcttacccttctgtgcagatcccaagct	5451
Db	2959	tttggagaggtatggtgttgagtaaatcaagaactcttaaccctgttagacaagtataatta	3018
QY	5452	gactccacaaaggtcctatccacagatcaagtatgtggaacgctactttgtatccctaag	5511
Db	3019	gattccctaaacaggtatatacttcagatatacttatgtgtgagccatactttgacacataag	3078
QY	5512	ctcaagagccgggtgacctactcttgaccggaactatggtgcttcgcacattccctgtctgc	5572
Db	3079	atgaaagagccagaacacacttatcttcgaaaaaatltaacaactcttcgttcgattctgtaacgt	3138
QY	5572	acgcgcgttcaacgcggatctgggtgcgcgcacaggggagctgcgcggaagaacacagttatg	5631
Db	3139	aacaccttacttaagtatgcgttcccatagggtgaacttcatagaacaatatcaagaagag	3198
QY	5632	acgcgtctcagacccgcagccacgcctccctccatcacaaagactcgcactccgtgtgtgcac	5691
Db	3199	accattctgactacgtctctatgctcttccattatataaacaaggtcaatgttccatcat	3258
QY	5692	cggtgagaggaacggtgctcgaaagccagtggtgtgtccatctgaagatcatgtgaagaagaaca	5751
Db	3259	aaagaagaggtatcatcttaacaaccaatgtgaagtgtctattgtgagacatgtgcagaaaaaagca	3318
QY	5752	cggtgagctgcgcttttgcacacgcagagacccacccagatgtctcaagatgtcaatagaatgtg	5811
Db	3319	cagagaggttgatcttgcaaacacatcatcgatcccgcgaccccaaaaatgcttcatagtgta	3378
QY	5812	cttcagagctctgttagggccacacggtgaacacaggtctcccttggaggtgcccagatgttt	5871

```
Db 3379 ctccagagatctgtacgaccacagatcaggggccttggaaagtgtgccaggtttt 3438
Qy 5872 tttagcagagatcccggaagaccacagctcttcgcgcacacaaatgtgcgctcgc 5931
Db 3439 ctgtctgaataactactgtgtacccaagctcttcacagacacataaactgtgactctgc 3498
Qy 5932 ttcaaggaactctgcaagaatgtgagatgtgcgtcgcggaataaataaggccctgattgg 5991
Db 3499 tttaagatttctactaaagtgatgtgaaagtgtccttaagaanaataagagcttaattgg 3558
Qy 5992 cgggaccaggaaggaagacacgcgtgagcgagcgaactactgtccgccttcgcggaggtc 6051
Db 3559 ccggttcaaaaggaatcacaagaaggaaatcg--gggaaactatcttcgccttaagaagacc 3616
Qy 6052 ctgcagccctcgtcttacccagcgccgcgcacgtgatgtgacaccacccc---accgcg 6107
Db 3617 ctacagccctagctacacagaaagtcctcagatcatccaagcaagatttgtctgcctcgcg 3676
Qy 6108 cctcaggaactccttgaacagagcaagttccgaaagcagacctgtgacg 6158
Db 3677 ccacagagatctcttcagtcgaatgagcttccgcaaatgagatctcttaacg 3727
```

RESULT 7

AAH41912 ID AAH41912 standard; cDNA; 4144 BP.

AC AAH41912;

DT 30-AUG-2001 (first entry)

DE Human CIASP-3 cDNA sequence Fig 4.

KW Human; cadherin-like asymmetry protein; CIASP; CIASP-3; Immune response;
cell surface molecule; transmembrane protein; immunosuppressive; vaccine;
antiinflammatory; antirheumatic; antianemic; dermatological; uropathic;
KW ophthalmological; antipneumatic; nephrotropic; antithyroid; antidiabetic;
KW neuroprotective; antisthmatic; antibacterial; antisense therapy;
KW gene therapy; chromosome 1p31.1; ss.

XX Homo sapiens.

XX MO200142297-A2.

XX PD 14-JUN-2001.

XX PF 13-DEC-2000; 2000WO-US34171.

XX PR 13-DEC-1999; 99US-0170453.

XX PR 14-JAN-2000; 2000US-0176195.

XX PR 14-FEB-2000; 2000US-0182296.

XX PR 11-APR-2000; 2000US-0196267.

XX PR 11-APR-2000; 2000US-0196460.

XX PR 11-APR-2000; 2000US-0196527.

XX PR 11-APR-2000; 2000US-0196528.

XX PR 11-APR-2000; 2000US-0547276.

XX PR 13-OCT-2000; 2000US-0240503.

XX PR 13-OCT-2000; 2000US-0240508.

XX PA (ARBO-) ARBOR VITA CORP.

XX PI Lu P, Garman JD, Candia AF;

XX DR WPI; 2001-375003/39.

XX PT P-PSDB; AAB99494.

XX PT Novel isolated cadherin-like asymmetry protein (CIASP) -3, useful for
autoimmune disease -
XX Disclosure; Fig 4A; 189pp; English.

CC The present invention describes an isolated polypeptide (I) comprising a
CC nucleotide (nt) sequence (S1a) that has at least 90% identity to the
CC sequence given in AAH41934 and is immunologically cross-reactive with
CC the derived amino acid (aa) sequence (S1b) given in AAB99495 or shares a
CC biological function with native CIASP-3. (II) has immunosuppressive,
CC antiinflammatory, antirheumatic, antianemic, dermatological, uropathic,
CC ophthalmological, antipneumatic, nephrotropic, antithyroid, antidiabetic,
CC neuroprotective, antisthmatic and antibacterial activities, and can be
CC used in antisense therapy, vaccine production and gene therapy. CIASP-3
CC related sequences can be used in preventing or treating a CIASP-3
CC mediated disease, preferably an autoimmune disease by inhibiting an
CC immune response. The autoimmune disease is caused or exacerbated by
CC increased activity of T helper 1 lymphocytes (TH1). AAH41933 to AAH41953
CC and AAB99491 to AAB99507 represent sequences which are used in the
CC exemplification of the present invention. CIASP-3 is localised in the
CC chromosome location 1p31.1.

CC Sequence 4144 BP; 1252 A; 829 C; 890 G; 1173 T; 0 other.

Query Match 22.7%; Score 1446.6; DB 22; Length 4144;
Best Local Similarity 64.5%; Pred. No. 6.8e-283;
Matches 2239; Conservative 0; Mismatches 1199; Indels 33; Gaps 4;

```
Qy 2713 cgcataccgtgcccagaagctgtctcaagagagctgtgctgcagctgtgtgtcagcagc 2772
Db 266 cgttaccacaactaaataaaagcttttccacagagagctgtgctgcagctgtgtgttgaagt 325
Qy 2773 agtgcgtacgcagggccatccctccagacgcctgtgtctctccagcagcattgtgtaag 2832
Db 326 ggcagcgcttgcggatcatcagcttgcacaagaagcctgttcttttgattaaatgtgtaag 385
Qy 2833 agtatgcgtcagactgtcgtcttgccagcagactgaacacaccccgcaagctgtgcctc 2892
Db 386 agcatggtgcacacatttatacttaactatgataaactgaggtcccaagaaagtcgcttt 445
Qy 2893 cccgagaccttccggaagacatacagctctgtgtggtctgtgtgcgcgcgaggtatc 2952
Db 446 ccagaacgttcaatggaatgacatgtgcaactctctgtgcagcaagatgtcagtatgataatgt 505
Qy 2953 accgtgtccacaagatgtgagctgtgcgcagacactcaacgccaagcctgtctctc 3012
Db 506 tcacgatttcagaagagacagaagaatgttgtagagactcaataaagcctgtgactctt 565
Qy 3013 ctcaagtaccttctgctccctgtgtgacccggcttgtcttcagcctgtgtccggccac 3072
Db 566 ctcaatgactgtgtgtctgtttagacagagattgtttttagcctataaagctctgc 625
Qy 3073 tacaagcaggtgtgcccagcgtctccagtcgctcccttaatccagcagcctgtgtgacctg 3132
Db 626 tataaacaggtgtcttcaaaagcttactacattaccagatccagtgctgtgtgtcttg 685
Qy 3133 cgcattgaaatccacccgactctgtgcagccagcagactagtcgtaacccctcaactccc 3192
Db 686 aggcgtgatttctcagatcatctgcagctatgacactatgttaataaacttaacc 745
Qy 3193 tgcgtccctctgtcaactcagcctgcctccctctgtgtcttcacaccatcccaag 3252
Db 746 tgcagcttacttactccacgtcgtcatctccacacactctgttcttctgtgaacatccag 805
Qy 3253 agctccacctctccagcagcccggaagcctgtgacccagctgacagcagctgagt 3312
Db 806 agtctgatttctcagaagatgtaacaagaatgtaacaataatgttgaattatcc 865
Qy 3313 ggaccattccgagcagcagcacttccatgctgtgtcctgtcgtgacgagctgtgcactggcc 3372
Db 866 gtgccttccgcacaacagcatatttgcagagacttgttaaacgagcgtgcgtcatc 925
Qy 3373 ctgacactgaggtgtgaaggagcttctgtgtgcacaagaagcctcagctgtgtcac 3432
Db 926 ttgaacctgtatgtcgaagagcttgttgattgcataagaagacatcatatgttacac 985
Qy 3433 agcctgctatgtgccaatgacatgcagcccgctcagcgcgagcactgtgaagcgctcgt 3492
```

D	986	aattactctccagtcacagctcgaagcccggtactctcgacctcaagataaagctcga	1045
Q	3493	gtgcccagctctgctcccgcaactgctctcgattgcacgggataaccttgccacgctgcgat	3552
D	1046	gtgcacatgtgtatctacacctctgattgattatcatgaacctgtacctcagctgatt	1105
Q	3553	gactctgttgaggcccaaggttcagcggctcaagctctgcttcattgttaccagaaca	3612
D	1106	gattttcaagaactccaatcaacacagaggaagac-----caattgtgtaagccctgat	1159
Q	3613	gaagcgcaagaggacattctgggtggtatccataaacccctctgtggccatgccaattgctg	3672
D	1160	gattatggaagtgagcgggaagtatagataagcgaacccgtgtgcattgccaatgcgaagg	1219
Q	3673	ggccccctgagccctctgctcccgggccagcatctccagggggccacaaagggcttcctgc	3732
D	1220	acatcgtgcccttcacatacaagaagcctgtgagttctctccatcagtcacaggtgcaag	1279
Q	3733	gcaggctgtgcccctctctgtctggttcaagccggaaacctctgtcggtgtgtgtgtggtg	3792
D	1280	caacacctacaccttttcagagataatcaatcagtcgaagccttttgatgcttacttctgggt	1339
Q	3793	ctgaaaaaacacacggccggcgcccccgcacagcgtctggcgcaacgactgacactccccaag	3852
D	1340	ctcaaaaatgagatggaataaagttctcaacgaagtggtttacaagatcctccagctcttcag	1399
Q	3853	ctggagacgtctgttggactctgtctgtacaccttgctgtgcctctgtgatacaagggaa	3912
D	1400	ctaaacgcgtattagctctgtcttactctctgtgtgtctgtgtgtgtgtgtgtgtgtgtgt	1459
Q	3913	aagccctttgaacgatacaacagcctccacaattcaaaaaactctctgtgatatgaagggcgg	3972
D	1460	aaagtgttttaacgaaatgaataagctgtgaccttataagaatacaaaagacatagagacaag	1519
Q	3973	ctagaggaaagacatctctgggtgtaccatccgagcgtccgcaagaaatgttccggcgaagtcgt	4032
D	1520	cttgaagaagactattctctgtggagcacaaggtgcacaggcaagaatgtgtctggcggaaccca	1579
Q	4033	g-----agaggagcccggtctgtgggaataccagagaaagctgcgtctgg	4071
D	1580	ggacagctcgcagaggaagcccatctggaagtgtccttgggaagtcagaagaatttgtagtgg	1639
Q	4072	cgggaaggcgctcaacacacactgtggaagcaaacctcaagacccggtgtgacagaaccaaagatga	4131
D	1640	agggaagaatatgtctccactctgtcgcttcaaaaacagagaagcgttgcacaataccaagagcagaag	1699
Q	4132	atggaacacagagccctgtgtggagaggaaacccggcgcaacccggagcagaagctgtgtctgt	4191
D	1700	attgaaacacaggaactgattgtatgtgaaacccggcgctacagaagcaacacttaatttta	1759
Q	4192	gacacactgtagatactactgtgcaagcgggtgatacttcaagaagcccggtgagagcgtctgt	4251
D	1760	gatacatatgaagattgttgttcaagaccggttctctgtaagaagataccaagaagagatctctc	1819
Q	4252	ggggcagtgctgtgaaggttgtgtgtgtacagcctgggcaggtgccagaagtgcctctctgt	4311
D	1820	gggtgtgaggtctaaagaatgtctacttaccacagacatgctgttlaaccaaagtcagttattcta	1879
Q	4312	cagagatgacctgcccacccagaggcgccctgtgtccaaagtcccgagctcgtgttcgag	4371
D	1880	caaacactgttctgtctacaaagagagcgtctgttttcaaaagttctctgtacactttaattgaa	1939
Q	4372	gaggaacagagtgctgtgtgcgacacccgtgtccggaggcccttcaagacatgtgtggcagccgc	4431
D	1940	gaagagagacagagagtgctgtatattatgtcctcaagctctctccgacacatgtaagcagtagc	1999
Q	4432	atcaagcaacatccgcagcagcagccagcgcctcgtgtatacctgtcaibtgcaacgaatcc	4491
D	2000	atcgtgtaacatacgtgcacacccacagtgctctcccttaccataatgtaggcaaaacttc	2059
Q	4492	gaagatcgcccaacttctgcggtgtgaagatgcaggatcacacgtctctctctgtccctg	4551

Dh	2060	ggaattgggaataaacttctgcgaagggttaaaatgcaggtaccaatgtcactatccctcttg	2119
Qy	4552	gttgggagacgacgcgaataacttcactgtaagagagacactcgcagatctcaactcaaaaccatctcc	4611
Dh	2120	tttgggacatctcagaatttaattgaagaattcttaagagcttctctaaagctatattg	2179
Qy	4612	acctatgctcgaagagagacatgggcctcgcggagacagcacctctcgacagagccaggccaggac	4671
Dh	2180	acaatgtcgtagaagaagactctgtaacttgaaggaaacaacacttccctgatcaagctccagat	2239
Qy	4672	ctgaattctaaacctcgcacatgatcttccttcgaggaacggtgaaggttgaaggaacaacagag	4731
Dh	2240	ctgggtttcaatctccatcccatatgatctcttccctgcgaataccgttgaaatgtgaagaaaccaagag	2299
Qy	4732	gaaccttgagatgctcatccgaacctcatgtacagaatttcgcccgggtgctacaagagctccacg	4791
Dh	2300	gattcctgtaaatgttgattgatcttaattgtacagaatttcgcaaggtttacagagctctcca	2359
Qy	4792	gaaccttcggctcgacctctgtcttgcaagaacatggccgggaagacgvcggagcttggccaacc	4851
Dh	2360	gactctgcgaattgacctcgtgttcgaaacaacatgycagacgacatccacacgaagaacatcat	2419
Qy	4852	ggcagagcgccgcgcagctgcatgttgctacgcgcgcgcctcttgctcgtgagtaactcgccctg	4911
Dh	2420	gcttgaagcttgcaacagttgtcctgatccactccagcagccctgtgtgcgaataattgtagcgt	2479
Qy	4912	ctcgaagagccagcgccacactcgccttgccgttgagcttgcttccctccagaacaatccatccac	4971
Dh	2480	ctggagagcccggaataatctctccgttgaggcttgtaacaatttcagatatttcaatctaat	2539
Qy	4972	gtgctagagaggtccgcgcactccgcagacacccctgtcgcccgacgagagagcttcttc	5031
Dh	2540	gtttttagaagaatcgtcgtctccagatgtagtgttctccagatgaaagagtatctgc	2599
Qy	5032	tcccggaagacacttcacactcagacctcgggctctgttagagttcttgtaacagagcgccgcgtac	5091
Dh	2600	ctctggaataatctcttaccagatccagacctctgttgagcttactcgtgaacaagcagcgtctcc	2659
Qy	5092	ttcaacacatvggctcggcgtcttcaacgagcgttgtaatgtagtcttaacaagaacatccatcccatc	5151
Dh	2660	ttctctatgctgtgcgtatgtatgaagcagttaatgaagtttacaagaatacttatcttcctatc	2719
Qy	5152	ctggaagcccaacgctgtgactaagaagaactcgtgcgcggtgacgcgcaaacctcagagagcc	5211
Dh	2720	catgaagcttaatctgggagtgaagaagaactatccacacatctcagttaaaccttcaagaagca	2779
Qy	5212	ttcaccaagagatcagacacccagagttccgcgtctggagcgcggtgttcoggagctatcttcgc	5271
Dh	2780	ttcaacaaaatgtgtctacacgaagtactcggcttggagacggaatgttttggaacatttccgt	2839
Qy	5272	gtgggctcttcaacgcgcacccactccggttgacactgtagtaagaagagcttcgtgtacaagag	5331
Dh	2840	gttggtttttatgtgaaccacaagttccgggatttggatgaagaagaattgtgttacaagag	2899
Qy	5332	ccatcgtatcacgaagcttgcgacagatctcaacccggtcgtgaagagttctacaagagaga	5391
Dh	2900	cttggaataacaaacttcgacagatatactccaagaattggagggattttacggaaga	2959
Qy	5392	tttggcgagcagcgtcgttggatattatcaaaagctcttaccctcgttggaacaagtccaagct	5451
Dh	2960	tttggagaggtgttgggtttgaagtaatcaaaagctcttaatccctgttgaacaagtgttaata	3019
Qy	5452	gactcaacaagaagcctacatccacagatccagattgtggaacgctacttgaacttaacgag	5511
Dh	3020	gattcctaacaagcatattatcgaattacctcttgggggcacataactttgacaataagag	3079
Qy	5512	ctcaacagccgggttgactactcttgcgcgcaactatgggcttgcacatcttcttctgc	5571
Dh	3080	atggaagcagaatcaccattatttcgcaaaaattacaactcttcgttgattcaatgtaactgt	3139
Qy	5572	acgcgcttcaacgcgcgatgtggcgcgcaacggggagcgtcccgagcaacaacagcgttaag	5631
Dh	3140	acacaccttactttagatggccgttgcacatggggaaattcttgaacattcaatcaagaagaa	3199

QY	3253	agctcaacacctctccagccagaagcccccagaccacaagtgacagcatgttgcgaactgaatc	3312
Db	861	agttcttgatttttttcttaacgaatgtatcaagaagacaaaagaatgtccaataattgttaattaccc	920
QY	3313	ggaccattccggagcgagacaccttccctgaagcgtgcctctctctgaagagctgtgcatctgccc	3372
Db	921	gtgcctttccgcgaacagcatattatgtgcagagactgtgtgttaacagagatgtgctgtcatc	980
QY	3373	ctcgaaaccttgaggtcttgaaaggggacattccctgtctcaaaaagaagccatcagttctgtgac	3432
Db	981	ttagacccttgatgtctgaagagactgtttgtgtgttcataagaagaatcatcataatgttaaac	1040
QY	3433	agcctgtcatgttgcgcattgcatactgaaccccgctacgcgcagagccatgttgaaagctcgt	3492
Db	1041	aatttactctccagttccacgcactctacgaccgcggttactctgcacccctcaataaagcttcga	1100
QY	3493	gtggcgcgagctgtacccgcgcacacgtccttcgatgtgacagagatcactcttgcacgctgat	3552
Db	1101	gtgcccattgtgttacttcaactctcgatgtgtattatatacgaagaaactgtctaccctgaagctat	1160
QY	3553	gaatttctcgagggcccgaagttcagtcggttcagagactgtgcctcaactgtgcatacgaacaca	3612
Db	1161	gattttcaagaactccaacatacagaaggaagc-----caattgtatagccactgtat	1214
QY	3613	gaagagcgaaagggagcatgtgcgggttacatcaaacccctctgtgtgcacatgccaatgtgtgt	3672
Db	1215	gattatgaagaatgtgaagcgggaagatgtatgaagaacgaacccggttgcattgtgcaatgcagag	1274
QY	3673	ggccccctaaagccccctgtctcccggtgcgaacatctccagaggggcccaacagggctctccgc	3732
Db	1275	acatcgttccctcaacatacaaaaagacctgtgcagtttccctctcagttcaacagatgtgcag	1334
QY	3733	gcagagctgtgccccctctctctgtctgtgaatcaagccggaacctgtgtgcgtgtgtctgtgagtg	3792
Db	1335	caaacacactcccttttccagagataagtcgaagccttttgatctgtcctactttgtgttc	1394
QY	3793	ctgaaaaaacccgagccgggcgcctctgcagagcgtctgtgcacatgcactgaacacatcccccag	3852
Db	1395	ctccaaaaatgcagaatgaaacagattccacagaagatgtgtttacagatctctccagcttccga	1454
QY	3853	ctggagacgtctgtgtgacatgtgcgtacaccttccctgcgtcccttggattaaaggggaaa	3912
Db	1455	ctaaacgcgtattatgaatcgccttatactctctgtgtgtcttgccttggagtataaaagggaaa	1514
QY	3913	aaaggcctttgaacgcatacaacagcctcacatcaaaaatactctgatatgaatgaagcgcg	3972
Db	1515	aaagtgtttgaacgaatgaaatagctgtgacctttaagaatacaaaagacatagaagcagaag	1574
QY	3973	cttagaagaagaacacattctcgtgttacacatccggaagcttcgaagaagaatgtgttcgycgaagctc	4032
Db	1575	cttgaaagaagctactctcttggaacataaagttgcacagcaagaataatgtatccgycgaagccga	1634
QY	4033	g-----agaaagacccgctttgggaatccagaaacgctgcgtcg	4071
Db	1635	ggacagcttcgagagaagagccocatctgtgaagtgtccttggaaagtccaagaataatttgaagtgtg	1694
QY	4072	cggagaagcgttcaacacacttggaaagcaaaccttcagaacccgctgtgaagaagaccaagatgaa	4131
Db	1695	aggaagaatattgtactcaactgtgcgttcaaaaacacagaagaagcttgcgcaaaatcaagcgagag	1754
QY	4132	atggaacaagaagcctctgtgtgaagaaggaaacctgcgaaccggagcgaacgactgtgttctgt	4191
Db	1755	atgtgaacaacgaacactgtattgtatggaacaccttgggtcttaagaagaacacactatcatctta	1814
QY	4192	gacacacttgaaagatcatctgtcgaagacgggtgaatgtcttcagaagaccgggaagagctgttg	4251
Db	1815	gatacatatgagagcttctgttcttcagacgctttctgttaacggaatcccaagaagagatctctc	1874
QY	4252	ggggcgagtgctgaaggtctgtgtctgtacagcctgtggcgagttccgacagatgtccctctctgt	4311
Db	1875	gtgtgagatgtctaaagtgctactatacagacgacatgctgttlaacaaagtgcagtttatcta	1934
QY	4312	cagcatgtgcctgtccacccagagggcccttgttccaaagttcccggaactgtgtgttcgag	4371

Db	1935	caacacgtttctgctcaacagagagccttctgttccaagtcttcctgaactcttatttgaa	1994
Qy	4372	gaagagacggagagctgtgtgtccgagcctgtgctcgttaaggctccctaaacactgtgtcagccg	4431
Db	1995	gaagagacagagagagtggtgtgttttatgtctccaggctctctcgaacactgtagcagtagc	2054
Qy	4432	atcaagcacatccgcgcagcagccagcgcctgcgtctaacctgtcatctgcagagaacttc	4491
Db	2055	atcgtgtacaataagtcgacacccccagctgctcccttactactaataatgagcaaaacttt	2114
Qy	4492	gaagatcgccacaactctgcgcgtgtgaaagatgcaggttacaactatgtctctctgcctgt	4551
Db	2115	gagattgggaataactcttgccaggtcttaaaatgcaagtccaatgtccactatccctctgt	2174
Qy	4552	gttggtgcgagcgcagaacttcagttgaagagcaacttgcagcgttactacccaacacatcctc	4611
Db	2175	gttggtgcacatctcaagaattttaatgaagaattcttaagaagcgtctctcctaagaagataatgt	2234
Qy	4612	acctatgctggagagagacatctgggcgtcgggacagacactctcgaaagaggttccagac	4671
Db	2235	acataagctgaagaagaactcgtgaattgagggaaacaacatctccgtatcagatccagatc	2294
Qy	4672	ctgaagtctcaacctgtgacatatgtactctgaagacacggtgaagaatgaaggaacacagag	4731
Db	2295	ctggtttcaatctccatgatattcttctctgatactgtgaanaatgaaaggaacacagag	2354
Qy	4732	gaacctgagatgtctcatgaacctcatgtacagaattgccccggggtccacagggctccacg	4791
Db	2355	gataccggaatgtctgattgtactcctaagtacaagaatctgcgaaggttacaagacactccaa	2414
Qy	4792	gaaccttcgctcgaactgtgtctgacagaacaatgcgggaaagcaacgcgtgtgacaaacac	4851
Db	2415	gatactcgatctgaacctgtgtcgaagaacatctgcaggaagcaagcactcaagaacgaagcaatcat	2474
Qy	4852	gcccagagccgcacagtcagttgtgcacgagcgccgcctctgctgtcgtagtaactgcctgt	4911
Db	2475	gctgaagctgcgcacagtgctctagttccaactaagaagaactgtgtcgtgaattatttgacatgt	2534
Qy	4912	ctcgaagacacaaacgcacacctgcgcgtggtgcgtctgtctcttcacagaacatccacacaa	4971
Db	2535	cctgagagcagcgaataatctctcctcgtgtggatgtgttaaaccttcagataattcactaat	2594
Qy	4972	gtgtctagagagatccgcacatctccgaagacatccctgtgcgccgacgagaggtctctgc	5031
Db	2595	gctttaagaagaatctctgcgtctccacaatgtatgtgtatctccagaatgaagaagtatctgc	2654
Qy	5032	tcocggaaacacatctcaactcgtcagctctgggcgtgtgaggttctctggaaacagcagccgcac	5091
Db	2655	ctctgnaaaatacttacttagtcagtcaggaactctgtggatatacttgaacaaagcagctctcc	2714
Qy	5092	ttcacacatgggcgggctctaagagcggtgaatgaggtctacaagaacctcatcccatc	5151
Db	2715	ttctctatgtcgtgcatagtatgaagaacagtttaatgaaagtttacaagtacttacttctat	2774
Qy	5152	ctggagagcccaacgtgtgacacaaagaagcgtgcgcgggtgacaggaagaactgaagagcc	5211
Db	2775	catgaagctaacccggatgtcaaaagaactatccacaatucacatgacaaacttcaagaagaca	2834
Qy	5212	ttcaccaagatcatgtcaccaaggtctccggtctggagacggtgtcttgaggacgttattccgc	5271
Db	2835	ttcagcaaaatgtctcatcagagtagtactcgtcgtggagcggaatgttgcacatatttctgt	2894
Qy	5272	gtggagctcttaagcgcgccacactgtgtgtaactcgtgatgtggagcagagattgtgtacaagag	5331
Db	2895	gttgttttcttaatgagacaagaattctggggatattgtgaatgaacaagaattgtttacaagag	2954
Qy	5332	ccatcatcatcagaacgtcgcacaagatctccacaacccgcgtggagggagtttacaagagaga	5391
Db	2955	cctgtgaataacaaactctgcagagatatactccacagatvggagggatttacaagaaaga	3014
Qy	5392	tttgcgacagagtgctgttgagattacaagaacttcaacctgtgtgaaagtccaagct	5451

Db	3015	tttggagagatgtgtgttgaagatcaaaaagactaatcctgtatgacaaagtgtaaalta	3074
QY	5452	gactcacaanaaggcctcaatccatccagatcaagtalgttgaacacgttacttggataccacag	5511
Db	3075	gactctaacaaggcatatattcaagattactatgttgagacataactttgacacatatgag	3134
QY	5512	ctcaaggaacgggtgtgacctactttgacgcgaactatlggtcttcgcacattccctgttcgc	5571
Db	3135	atgaaagagacagaaatcacctctatttcgacaaaataattcaaatcttcgtcatctatgactgt	3194
QY	5572	acgcgcgttccagcgcgagttggcgcggaacacgggagctgtgccgagcaacacaagaugtga	5631
Db	3195	acaccccttactttaagttgvcgctgcgcctcctcgaatctcatgatacaatcaaaaagagag	3254
QY	5632	acgctgtctcagcacgcgacgcgctccctccatcaataagactcgcgtatgtgtgcac	5691
Db	3255	accattctgactaagctctaatgtccttctcttatatataaacaagtgtaattgtccattat	3314
QY	5692	cggagagagacggtgtctgacgcgcagctggaagtgtgcatcgaaagacatgcagaagaaga	5751
Db	3315	aaaagaagagatcatcttaaacacaaatttgaagtgtcatgtatgagacatgcagaanaagaca	3374
QY	5752	cggagagctgtgccttttgcacacgagcagggcccaacagatgctcaagatgtctacagatgtt	5811
Db	3375	caggagctgtgcattctgcaaacacacacagatcccgacagaccacaaatgtcttaagttgta	3434
QY	5812	cttaagggctctcttgaaggccacacgctgaaacagagttcccttgaaagtgtggccagatgtt	5871
Db	3435	ctccagggatctgttaggcacacacagtgaaatcgaaggccctttggaagttgccaggtttt	3494
QY	5872	ttaagagagatcccggaagaccacaaagctcttcgcgcatcacacaanaatltgcgctctgc	5931
Db	3495	ctgtcttaaatactagttgagcccaaaagcctcttaagacatcataataaactgagactctgc	3554
QY	5932	ttaaggaactcttgcgaagaaattgtgagctgtgcgtgcggaataaataagaccctgattgg	5991
Db	3555	tttaagaattttacttaaaaagtgtgtgaagtgtccttaagaaaaataaagaagcttaatttgg	3614
QY	5992	ccggaaccaaagaagatgaacacgcgtgagcctgagacgaactatgcgcgtgcggaagct	6051
Db	3615	ccggttcaaaaagagatcaaaaggggaattg--ggbaactatcttcgccttaaaagagcc	3672
QY	6052	ctgcagcccccgtcttaaccaagcgcgtgccccaagctgtatgacacccaacccc---accgg	6107
Db	3673	ctacagcccttaagatcaacaagaagttccctcaagtatccaagccagattgtctgtccctg	3732
QY	6108	ccctaggaactcctcttgacaacagagcaagttctcgaaagcagagcctctgagc	6158
Db	3733	ccaaagagatctcttcagttcgaatgtgacttgcgcaaaatgagatctctctaaac	3783

RESULT	9	
AAH41952		
ID	AAH41952	standard; DNA; 3173 BP.
XX		
AC	AAH41952;	
XX		
DT	30-AUG-2001	(first entry)
XX		
DE	CLASP-3 nucleotide fragment.	
XX		
KW	Human; cadherin-like asymmetry protein; CLASP; CLASP-3; immune response;	
KW	cell surface molecule; transmembrane protein; immunosuppressive; vaccine	
KW	antiinflammatory; antiarthritic; antianemic; dermatological; uropathic;	
KW	ophtalmological; antirheumatic; nephrotropic; antithyroid; antidiabetic;	
KW	neuroprotective; antislumatic; antibacterial; antisense therapy;	
KW	gene therapy; chromosome 1p31.1; ds.	
XX		
OS	Unidentified.	
XX		
PN	W0200142297-A2.	
XX		
DD	14-JUN-2001	

[illegible]

QY	3703	atctcccaaggaggccaaccaagggctcttcgcgcgagcgtctgtccctctctgcgtgattcaagc	3762
Dp	358	agttctccctccatcgttacaacgaatgtgcagagcaacacactacccttcttcagcagaattcaagt	417
QY	3763	cggacactctgtgcgcgtgtgtgcgtgtgcgtgcgtgaaataaacccagagccggcgtccctgaag	3822
Dp	418	cgaagccttttgatctcgtctactcttgggttcccaaaaatgcagatagaacaagttcttaag	477
QY	3823	cgctggcgcaatgcacgtgacactcccccagctcggatgcgtctgttgcactgtctaacctt	3882
Dp	478	aagtgtgtttacagatccctccatcgtctgcagcttaaacccggtctatgactctttatctc	537
QY	3883	tgcctgcgtctgccttttgatgatacaaggggaaaaggcctttgaagcatataaagcctca	3942
Dp	538	tgtgtctctgcctcttgatgtataaagggaanaaagtgtttgaaacgaaatgaatagcttgacc	597
QY	3943	tccaanaaatctcttgataatgaaagvgcgcgctctagaaggaagccattctgtgatactcgga	4002
Dp	598	tttaagaataccaagaagactagagagagaagaagccttgaaagagctattctctggagacatggt	657
QY	4003	gcttcgacagaatctgtgttcggcggaagtcgt-----agaaggac	4041
Dp	658	gcacggcaagaatctgtacgvcgaaagccgagacgctcgaagagaagccactctggaagt	717
QY	4042	ccggtttggagatccagagaagcgtgcgtctgcggaagagaggttaccactctgaaagcaaac	4101
Dp	718	gctcttggaagatccagaanaattctgaagvtgaaagaaagatatactaccactctgcgtacaac	777
QY	4102	tcaagaccgcgtgcgaagaccagaagatgaaatgaaacaaagagccctctgtgtgaagagac	4161
Dp	778	acaggagagagcttgtaacaatcagaagagatcttgaaacaaagacacttgattgtatgaaac	837
QY	4162	ctgggcaaccgagcgaaagcctagtgtgtcttcgcgacacactcggagatatactcgtgcagaagctg	4221
Dp	838	ctggctatacaagaagaaggaatcatattttagataatactagatgattgtgttccagacctt	897
QY	4222	atgctttccaagaagcccgvgagcgtctctgggggcgcgtctgaaggttgcgtctaacgc	4281
Dp	898	ctgttaacggaatcccaagaagagactctctgtgtgagatgctataaagtctactacaagc	957
QY	4282	ctgggcagctgcacagatgcgcctctctctgtgcagcatgtgcctgcgaccccaaggcgccct	4341
Dp	958	atggcctgtataccaagaatgtgagtttactatacaacctgttctgtctacaagagagccttg	1017
QY	4342	gtgtcccaagttcccgagcgtgcgtcttcgcggagggagacggagactgtgtgcgcgacctgtgc	4401
Dp	1018	gtttccaagtttcttgaactctatttgaagaagagacagacagctgtgcgtatttatagc	1077
QY	4402	ctggagcgtccctacgaacactgtgtgcagccgcatcagaacacatccgcagccgcagcgcc	4461
Dp	1078	ctcagcgtcttcgcgaacactgttagcagtagcatcgtgtacatatcgttccacccccaggtgc	1137
QY	4462	tgcgtgtaccctgtcatgtgcgaacgaacttcgagatccgcgcaaacacttggcccggttgaag	4521
Dp	1138	ttcccttactactaataagggcaaacatttggatgttgggaataactatgtccaggttaaa	1197
QY	4522	atgcaggttcaacatgtctctctcgttcccgtgtgggagcgaagcgaacttcagtgaagag	4581
Dp	1198	atgcaggttacaatgtctcatctactctctctgtgtgggcacatccagaaatttataatgaagaa	1257
QY	4582	caactgtgaggttcatcactcaaaaacacatccctcacacttctgtcagtgagagacatgggctgcg	4641
Dp	1258	ttcttcaagaagttctcttaagaagactatataatgacatatgtcgaagaagatcttggaaatggag	1317
QY	4642	gacacgacctctgcagagcaggtgcgcagggaccgtgatttcaacctgcacatgatctctgag	4701
Dp	1318	gaaacaacattctctgatcatcaggtccagggctcgtgttttcaactctcatatgatcttctct	1377
QY	4702	gacaacgttgaagatgaaagaaacaaagagaccctgtgaatgtctcatcgacctcatgtac	4761
Dp	1378	gatactctgaaatgtgaaggaacaccagggagatccctggaatgttgattgattctaatgtac	1437

QY	4762	agaattgcccgggagctaccagaggtccacgggaccttgctgagactggtttgcagaactg	4821
Db	1438	agaattgcccagaagggttaccagagctccctccacgactccgagcttgacctggtttgcagaactg	1497
QY	4822	gcccgggaagcagcgcgagacttgggcacacacgcccggagcgcccgacgtgacgtgtgcacg	4881
Db	1498	gaaggcgaagcacttcagaaacgaaacatcatgtctgaagctgcacagtgctctagctccatca	1557
QY	4882	gcccctctcgctgcgtcgtgatacctcgccctgtctcgagagccagcgccacctgcgcgtgggc	4941
Db	1558	gaagacactgtgtcgtgaatatttgcagctgcgtgagagccgggaaatattcttcgttccca	1617
QY	4942	tgcgttccctccacaaacatccatccacacgctgcacagagagttccgcacatcccgacgc	5001
Db	1618	tgtgtacaacttccagaaatatttcatctaatgttttaagaagatctcgctgcctcagatgat	1677
QY	5002	atccctgcgcgcgcagcagagaggtctctgcctccgggaaagcaacttcacgtgagctggcgctg	5061
Db	1678	gtgtatctcccgatcagaaagaggtatctgcgtctctggaaaaatctcttaccgtgactagagat	1737
QY	5062	gtagagtgctgtgcagacagcagcgacttccacttcacacatggcgcggtctatagagcggtg	5121
Db	1738	gtgggatacttcgtgaacacagcagctgcctctctcttaagctgtgcagtgtatgaagcggt	1797
QY	5122	aatgggtctacaagaacactcatcccatctcgtbgaagcccaacgctgtactacaagaagctg	5181
Db	1798	aatgaaagtccaagaagtacttacttctcttataagagctaatcggatcgtgacaagaacta	1857
QY	5182	gcccggcggtgcacggcgaacacttgagagagcccttccacaaagatcatgacacagaagttccgc	5241
Db	1858	tccacaacttcacaggtcaaaccttcaagaagcattccagacaaatgtgtcatccagagtaacgggc	1917
QY	5242	tggggcgcggtgttcgtggagcgtatctccgcgttgggtctctacgggcgccaaacttcggtgac	5301
Db	1918	tgggagcgagtgattgtgcacatttctcgtgtgtgtgtttttatgagaccaaagttcggggtat	1977
QY	5302	ctggagttgagcagagattgtgtgacagaagagccatcgatccacagaagcttgccgagaatctca	5361
Db	1978	tgggattggaacaagaattcgtttacagaagagcccgcaataaccaaacttgcgcgagatatact	2037
QY	5362	caacgcgtggagagagttctacacagagagagatttggcgacgagcgtctcgtgagatattacaa	5421
Db	2038	cacagatctggagggatttctacggaggaagaagatttggagggagtggtgttggaaglatcaaa	2097
QY	5422	gaccttaccctctgtgcagaaggtccaaagcttgatactacaaaaggcttaactccagatcag	5481
Db	2098	gaccttaactcctgtgacagaagtgtaaatltagatccctcaagaaggatataatctcagattacc	2157
QY	5482	tatgtggaaacgttaattgtgatacttaccagagctcagaagcgggtgtaccatactttgacgcg	5541
Db	2158	tatgtggagccatactatttgcacacatagtagatgaagagcagaatcaactatcttcgacaaa	2217
QY	5542	aaactatgggtctgcgacatctcgttctgcacgcccgttcaacgccgagatgggcgacac	5601
Db	2218	aaattcaactcttcgtgattctgattctgactgtacgttcaaccccttactttagatggcggtccat	2277
QY	5602	ggggagcgccccgagcaacacaaagtgtaagcgcgtcctcagcaacgcgaacgaccttccc	5661
Db	2278	gggggaacttcatagaaacatccaagaagagacatcttcgactacgcttccatagucttctct	2337
QY	5662	taccataagagctcgcatactcgtgtgtgcacacggggagagagcgtgtcgtacgcacagttgag	5721
Db	2338	tatatataaacaaggttcaatgtctcctcatbaaagaagagatcatcttatacccaattgaa	2397
QY	5722	gtggccatcgagagacatgcagaagaagacagcgagctgtgccttgcacacggagcaggag	5781
Db	2398	gtctctatgtggagacatgcagaaaaagcacaaggaagttgcatgtgcacacatcagagat	2457
QY	5782	ccaacagatgtcacaagatctcaacagttgttgcacacgggtgcctgtgagggccacacgttgaac	5841
Db	2458	cccgcagagccccaataatgcttcaagatgtgtactccagggagctctgtgagacacacgttgaat	2517
QY	5842	cagggctccacttggaagtgtgcaccaagttgtttttagacaagatccccgaagacccaagctc	5901

||||| ||| ||||| ||||| ||| | ||| ||| ||||| ||||| |||||
Db 2518 caggggccttggaaagttgcagcgttttctgtctgaataactagtgacccaagctc 2577
QY 5902 ttccgcgcacacacaatctgcgcctctgtctcaaggactctgcgaagaatgtgagat 5961
||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
Db 2578 ttgagacatcataataactgcgacctgtcttaagaattacttaagaagtgtgagat 2637
QY 5962 ggcgtcgcgaataataagccctgattggcgcgacagaagaagatgacacgcgtgagctg 6021
||| ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| ||| ||||| |||
Db 2638 gctttagaataataagacttaattggtccggttcaaaagagatataaagggaatg 2697
QY 6022 gaggcaacacacgcgcgcgcgtctgcagccctgttaccgagcgcctgagcc 6081
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2698 --gggaactatcttcgcttaagaagcgcctacagccttagatcacagaatgacctta 2755
QY 6082 cagctgattgacacccccc---accgcgcctcagaactccttgaaagagaattt 6137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2756 gtatcccaagcagatattgttgcctccctgcacagagatctccttcagtcagatgagctt 2815
QY 6138 ccgaagagcagactctgagc 6158
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2816 tcgcaaaatgagatctctcaac 2836

RESULT 10
AAS07382
ID AAS07382 standard; cDNA; 4027 BP.
XX
AC AAS07382;
XX
DT 26-SEP-2001 (first entry)
XX
DE Human DNA associated with CLASP-5 #2.
XX
KW Human; CLASP-5; cadherin-like asymmetry protein; immune gateway;
KW immunogen; antibody; autoimmune disease; rheumatoid arthritis;
KW multiple sclerosis; leukaemia; insulin dependent diabetes mellitus;
KW acquired immunodeficiency syndrome; AIDS; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2970
FT /tag= a
FT /product= "Protein as displayed in AAU04026"
FT /partial
FT /note= "No start codon"
XX
PD WO200142296-A2.
PN 14-JUN-2001.
XX
PE 13-DEC-2000; 2000NO-US34163.
XX
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
PR 11-APR-2000; 2000US-0196267.
PR 11-APR-2000; 2000US-0196460.
PR 11-APR-2000; 2000US-0196527.
PR 11-APR-2000; 2000US-0196528.
PR 11-APR-2000; 2000US-0547276.
PR 13-OCT-2000; 2000US-0240503.
PR 13-OCT-2000; 2000US-0240508.
PR 13-OCT-2000; 2000US-0240543.
PR 13-OCT-2000; 2000US-0240539.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PI Lu P, Garman JD, Candia AF;
XX
DR WPI: 2001-367865/38.
DR P-PSDB; AAU04026.

XX
PT CLASP-5 polynucleotides, proteins and antibodies are used to prevent or
PT treat a CLASP-5 mediated disease, such as an autoimmune disease e.g.
PT Rheumatoid arthritis
PS
XX
Dislosure: Fig 9C; 188bp; English.
CC The sequence encodes a protein presented as AAU04026, included in the
CC specification which relates to a cadherin-like asymmetry protein,
CC CLASP-5, which is a transmembrane protein of the immune system involved
CC in the formation of the immune gateway. CLASP-5, polynucleotides encoding
CC it and an anti-CLASP-5 antibody are used to prevent or treat a CLASP-5
CC mediated disease, such as an autoimmune disease caused or exacerbated by
CC increased activity of TH1 cells. These diseases may include Addison's
CC disease, Rheumatoid arthritis, dermatitis, Multiple Sclerosis, Systemic
CC lupus Erythematosus and Autoimmune Thyroiditis, inflammatory conditions
CC (e.g. Ischaemia-reperfusion) and responses, leukaemia, acquired
CC immunodeficiency syndrome, AIDS, syphilis, typhoid, Burkitt's lymphoma,
CC Grave's disease and insulin dependent diabetes mellitus. CLASP-5 is used
CC for inhibiting an immune response in a cell (T cell or B cell) by
CC interfering with the expression of a CLASP-5 gene in the cell, the
CC ability of a CLASP-5 protein to bind to another cell or the ability of a
CC CLASP-5 protein to bind to another protein. The antibody binding CLASP-5
CC is used to inhibit an immune response is a subject. The polynucleotide is
CC used to detect CLASP-5 expression in cells and for diagnosis of
CC diseases and disorders associated with aberrant expression of CLASP-5.
CC Note: The present sequence is included in the specification but is
CC not mentioned anywhere in the specification.
XX
SQ Sequence 4027 BP; 1162 A; 898 C; 957 G; 1010 T; 0 other;

Query Match 16.2%; Score 1031.8; DB 22; Length 4027;
Best Local Similarity 61.5%; Pred. No. 5.2e-199;
Matches 1771; Conservative 0; Mismatches 1072; Indels 36; Gaps 6;

QY 3218 cggccctcccccctgtgtctccaccacctccccaagctccactccagccaagccc 3277
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 ctccaacatctctgtctctccatctatctccgaactcaagctcctgcctccagcttc 82

QY 3278 cggaccccaagtgacacgacatgtgtgacatgtgtgacacat--ccggacacgacact 3335
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 aggacagaagaatcgccagcatgttcgactccagatccagctccgcgcgagcatt 142

QY 3336 cct-aggctggctcctgcgcagcagcgtgcgcctgcgcctcgaactgaagctgaagg 3394
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 cctcacggggctcctctcaccagacatgctgcgcctgcctgagatccgaaggagaa 202

QY 3395 catctctgtgcacaagaagcgcacatgtgtgtgcacagcctgctatgtgccaatga 3454
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 tcagcgaagtacaagaagaagctgtcagtgcaattcacagcctgtaagtctcagacc 262

QY 3455 ctgaccccccgaagcgcagcagcactgtgaagcgcgtgtgacgagtgtaacggcacc 3514
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 tggaccacagctgtgtccaacacagaggtgaaggtcaaaatccgccttaccactt 322

QY 3515 tgccttgcattgcagggatcctgtgcacagcgtcagctacttgctgagggccaggtc 3574
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 tagttgacatcatattgtgtgttggcaccagcctgtgacttaccagtgtg----- 373

QY 3575 agcgtgtcaagactgtgctcaatgtctgactcagacacagaagggcgaaggaattgtgg 3634
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 374 -----cagatactcgcagaataccgcacacagctgtctgagtgaagaagaagagcgc 427

QY 3635 gtaccatcaaccctctgtgcatgtgcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3694
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 gtgcattaccagaatgt 487

QY 3695 gggccagcatctccagagggccacacagcgttctgcgcagagctgtgtcctctctgtgt 3754
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 gtggaatagt-----gtctctctctgtccctataagcagtaacaacatgtcgaacgcgg 541

QY 3755 agtcaagccgagactgtgcgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3814

Db 738 atctgagatgctcgagagccacagctactgcgcgtggcgagtgctcaagcttcacgaata 797
Qy 4961 tctcatcaacgtgctagaggagctccgcatctcgacgaacatctgtgcccagagag 5020
Db 798 ttctctcaaatgctcgagagagctgtgtctctgaagacacccgtcaactcgagag 857
Qy 5021 agggcttcgtcccggaagcaacttcaactgagctgggtgtaagtgctgagcaag 5080
Db 858 atggggtgtgcgagagcccaacttcaacgagagtgctgtgagctctccggagcaag 917
Qy 5081 caagccgctacttaccatctgagcggtctctacgagcggttgaaatgagctacaagaacc 5140
Db 918 ccgagagctcttcagcacgagggcttatatgagacagttaatgagctactacaagcttg 977
Qy 5141 tcatcccatctcggaagcccaacgctgactacaagaagctgcccgcgtgacagcaac 5200
Db 978 tcatcccatctctagaaagcgcatcgagaattccggaagctgacacttcaactcaacgaagc 1037
Qy 5201 tgcagagagctctcaccagaatcatcaccagaagctcgcgttgagagcgctgttcggga 5260
Db 1038 tgcagagagctctcgcagacagcatcgttaacaagatcata-----agagaaigtgtgaa 1091
Qy 5261 cgtattccgctggtggtctctctacggtcccaactcgtgtgacbtgga tgcagagagtttg 5320
Db 1092 cctacttcgagttgttctcttgatccaatttggtgagttgtgatacagagagtttg 1151
Qy 5321 tgtacaagagagccatcgatcacaaagctgacagaatcttcaacacgcggtggaggtctc 5380
Db 1152 tctcaaaagagcctgcaaatcaacaaagctctctgaatcttcaactagaactagaagcattc 1211
Qy 5381 acacggaagaatttgcgcagacgctgtgtgagatlaacaaagactcttccctgttgaca 5440
Db 1212 atggtcaatgttttgtgtgcagaattgtgtgagtgatgaaagctcactctcctgttgaca 1271
Qy 5441 agtccaagcttgactcaaaaagcctacataccagatcagtatgtggacgcgtacttg 5500
Db 1272 aaacccaagtgtga tcttaaaaagcctacatacacagatcacttgtgtgagcctacttg 1331
Qy 5501 atacctagagctcaaggaagcggtgacctacttgaacgcgaacttggtgcgacat 5560
Db 1332 atgagtatgagatgaaagacaggttccacatactcttgagaagaattcaactcccgaggt 1391
Qy 5561 tctgttctgcacgcgcttcaacgcggtatggtgcgcacacagggagagctgcccagaaac 5620
Db 1392 tcatgtacacaccccggttcaaccttgagggggtcgtcgggagagctgcatgagcagt 1451
Qy 5621 acaagcgtlaagacgctgctcaagcaccgaacagccttccctacatccaagactgcacatc 5680
Db 1452 acagaaggaacacagctctgaacatactgacgccttccctacatccaagacagatcat 1511
Qy 5681 gttgtgcccacgaggaagacggtgctgacgcagtgtaggtggtgcacatcgagacatgc 5740
Db 1512 gctgtacatcagaagagaggttgtgttgacacagatgtaagttgcatlgaagacatga 1571
Qy 5741 aagaagaagacacgagagctgtgcttgcacacgagcaggaacccaacagatgtctaagatgc 5800
Db 1572 agagaagaagccctgcaggttagctgtgcattaaacagagcgctgtgtgcaagaatgc 1631
Qy 5801 taagagtggtgtctcaggggtctgttagggcccggtgaacagaggttccctggaggttg 5860
Db 1632 ttaagaagtgctgcaaggtctgtggaagctactgtlaatcaagggaacacatggaagtag 1691
Qy 5861 cccaagtgtttttagcagagatcccggaagacccaagctctcccggtcacatcaacaagt 5920
Db 1692 cccaagtgttttgtgcgaatctccgtgcatcacaactctatgcacatcaacaagt 1751
Qy 5921 tgcggtctgtctcaagactcttgcagaagaatgtgagatgctgctgcggaanaataag 5980
Db 1752 tgaagtatgtcttaaggaattcatcatgagatgtgtgtagtagaanaaacaagc 1811
Qy 5981 cctgattgtgcccgaagcagaagagtagacgcgtgagctggtggcgcaactactgcccgc 6040

Db 1812 gtctcatcagcgagacagaggaataatcagcaggaactcaaaaagactataaagc 1871
Qy 6041 tgcggagagctctgaagccctgttaccagagcgccgcccagctg 6087
Db 1872 taagaagaacctcagccaatgatgcagcggaataatccagaactg 1918

RESULT 12
AAC93501/c
ID AAC93501 standard; cDNA; 1273 BP.
XX
AC AAC93501;
XX
DT 16-FEB-2001 (first entry)
XX
DE Human secreted protein gene 23 SEQ ID NO:33.
XX
KW Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
KW antiproliferative; cytosolic; cardiant; vasotropic; cerebroprotective;
KW neurotropic; neuroprotective; antibacterial; vitruicide; fungicide;
KW ophthalmological; vulneryary; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorders; cancer; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
KW Alzheimer's disease; ocular disorder; wound healing; skin aging; ss.
OS Homo sapiens.
XX
PN M0200061626-A1.
PD 19-OCT-2000.
PF 06-APR-2000; 2000WO-US09066.
PR 09-APR-1999; 99US-0128698.
PR 20-JAN-2000; 2000US-0176926.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (ROSE/) ROSEN C A.
PI Rosen CA, Ruben SM, Komatsoulis G;
XX WPI: 2000-619227/59.
DR P-PSDB; AAB51849.
XX
PT New nucleic acid molecules encoding 49 human secreted proteins for
PT diagnosing, preventing or ameliorating medical conditions and used for
PT food additives or preservatives -
XX
PS Claim 1; Page 447; 516pp; English.
XX
CC Polynucleotide sequences AAC93479 - AAC93527 represent cDNA encoding
CC human secreted proteins AAB51827 - AAB51875. Sequences AAB51876 -
CC AAB51927 represent alternative polypeptides encoded by the genes, and
CC amino acid sequences with which they share homology. The genes and
CC proteins have activities dependent on the tissues and cells in which they
CC are expressed. Examples of their activities include: immunosuppressive;
CC antiarthritic; antirheumatic; antiproliferative; cytosolic; cardiant;
CC vasototropic; cerebroprotective; neurotropic; neuroprotective; antibacterial;
CC vitruicide; fungicide; ophthalmological; and vulneryary. The secreted
CC proteins, polynucleotides, antagonists and agonists may be useful in
CC treating, preventing and/or diagnosing diseases and disorders such as
CC autoimmune diseases e.g. Rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC infections caused by bacteria, viruses and fungi and ocular disorders
CC e.g. corneal infection. The polypeptides can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
CC carbohydrate, vitamins, minerals, cofactors and other nutritional

CC components. Oligonucleotides AAC93470 - AAC93478 and peptide AAB51826 are used in the isolation and characterisation of the proteins and CC polynucleotides of the invention.

xx Sequence 1273 BP; 254 A; 336 C; 366 G; 308 T; 9 other;

Query Match 13.0%; Score 829.4; DB 21; Length 1273;
Best Local Similarity 87.3%; Pred. No. 3.3e-158;
Matches 1005; Conservative 5; Mismatches 3; Indels 138; Gaps 4;

```
QY 5209 gcccttcacaaatcatgtacacagagttccggtctggagcggtgttctggagcgtatttc 5268
DB 1267 gcccttcacaaatcatgtacacagagttccggtctggagcggtgttctggagcgtatttc 1208
QY 5269 cgcgtgagcttctacagcgcccaactcgtgtgacccctgagtgagcaagagttgtgtacaag 5328
DB 1207 ccgctggcggtttctagcgcccccacattccgtgacacgtgagtgagcaagagttgtgtacaag 1148
QY 5329 gagccatcatcagaagctgtgacagagatctcacacggcgtgagagagttctacaagag 5388
DB 1147 gagccatcatcagaagctgtgacagagatctcacacggcgtgagagagttctacaagag 1088
QY 5389 agatttggagcagcagctgtgagatttcaaaagcttaccctgtgacagagctcaag 5448
DB 1087 agatttggagcagcagctgtgagatttcaaaagcttaccctgtgacagagctcaag 1028
QY 5449 ctgtactcacaagaagcctacacatccagatcagtatgtgaacgcgtacttctatacctac 5508
DB 1027 cttgactcacaagaagcctacacatccagatcagtatgtgaacgcgtacttctatacctac 968
QY 5509 gagctaaagagcgggtgtgaccttctgacccgaactaagggtctcgacatctctgttc 5568
DB 967 gagctaaagagcgggtgtgaccttctgacccgaactaagggtctcgacatctctgttc 908
QY 5569 tgcacgccttcaacgcggatgtgacgcacacggagagctcccgagcaacaacagcgt 5628
DB 907 tgcacgccttcaacgcggatgtgacgcacacggagagctcccgagcaacaacagcgt 848
QY 5629 aagaagctgc-tcaagcaccagacgccttccctacatcaagaactcgatcgtgtgtg 5687
DB 847 aagaagctgc-tcaagcaccagacgccttccctacatcaagaactcgatcgtgtgtg 788
QY 5688 ccaacggagagagagctgtcagccagctgtaggtgtgcatcagggagatcagagaaga 5747
DB 787 ccaacggagagagagctgtcagccagctgtaggtgtgcatcagggagatcagagaaga 757
QY 5748 gacacggagagctgtccttgcacacgagacagaccacagatgttaagatgtacaagat 5807
DB 758 ----- 757
QY 5808 ggttgctcagggtctgtagggccacacgtgaacagaggtcccttgagggtggcccaagt 5867
DB 758 -----TGAGAGTGGCCACAGT 743
QY 5868 gtttttagagaagatcccggaagaccacaagctcttcgggacatcaaaaatttgggt 5927
DB 742 gtttttagagaagatcccggaagaccacaagctcttcgggacatcaaaaatttgggt 683
QY 5928 ctgcttcaagagacttctgcaagaatgttagagatgcctgcgcggaataaagccctgtat 5987
DB 682 ctgcttcaagagacttctgcaagaatgttagagatgcctgcgcggaataaagccctgtat 623
QY 5988 tgggcccgaacagaagagtagtaccagctgtagctgtgagcgcaactactgcgcctcgga 6047
DB 622 tgggcccgaacagaagagtagtaccagctgtagctgtgagcgcaactactgcgcctcgga 563
QY 6048 ggtctgcagagcccttctaacacagcgttgcacagcgtgtagtggagccacccacccgg 6107
DB 562 ggtctgcagagcccttctaacacagcgttgcacagcgtgtagtggagccacccacccgg 503
QY 6108 cctcagaagactccttgaacagagcaagtttcgaaagagagacacttgcaccacaaga 6167
DB 6108 cctcagaagactccttgaacagagcaagtttcgaaagagagacacttgcaccacaaga 6167
```

```
DB 502 CCTCAGCAACTCCTTGACAGAGCAAGTTCGAAAGGACAGACCTGTGAGCCACACAAGA 443
QY 6168 ccaagctgtacacagagagacagacacccgggctcagctgtctgtcgtgagggag 6227
DB 442 CCANAAGCTGTACGTAGAGAGAACAGACACCCGGGCTCAGCTGTGTGCTGAGAGGAG 383
QY 6228 tctgcctgtgtcccaactggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 6287
DB 382 TCTGCTCTGG-GYCCACTGGGCTGTGGGTGACACACTGTACTTNGGCT-6GCCCTCT 325
QY 6288 gccctgtgtcccatctgtgtgacatgacttccctccctttaaattaaatgtgt 6347
DB 324 GCCCTGTGTGCCATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 265
QY 6348 ttataagca 6358
DB 264 TTTATTAAGCAA 254

RESULT 13
AAC75563
ID AAC75563 standard; cDNA; 586 BP.
XX
AC AAC75563;
XX
DT 08-FEB-2001 (first entry)
XX
XX Human ORF1118 polynucleotide sequence SFO ID NO:2235.
DE
XX Human: open reading frame; ORF; detection: cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; immunosuppressive; antidiabetic;
KW hypotensive; dermatological; coagulant; vasotropic; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200058473-A2.
PN
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000MO-US08621.
PE
XX
XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
XX (CUBA-) CURAGEN CORP.
PA
XX
XX Shinketsu RA, Leach M;
PI
XX
XX WPI: 2000-602362/57.
DR
XX P-PSDB: AAB41354.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PI neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 1649; 5507pp; English.
XX
XX AACT74446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORF open reading frames 1 to 3161. The ORF
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
```

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antihaltic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation, to inhibit thrombosis, and as a contraceptive.
 XX
 SQ Sequence 586 BP; 122 A; 185 C; 156 G; 123 T; 0 other;

Query Match 7.4%; Score 468.6; DB 21; Length 586;
 Best Local Similarity 98.0%; Pred. No. 2.5e-85;
 Matches 465; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 519 gaatgactcccgagctgctgagctcccggaagacacccctcgaagcagtgctcctc 578
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 88 gaatgactcccgagctgctgagctcccggaagacacccctcgaagcagtgctcctc 147
 QY 579 tagcattctcgactgaagaaacttgagcactactactatgtgctcctctcgtagacg 638
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 148 tagcattctcgactgaagaaacttgagcactactactatgtgctcctctcgtagacg 207
 QY 639 ggcggccccaagaagatgtgacggcgcaatgaacaccttgacggcagcagccgcccc 698
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 208 ggcggccccaagaagatgtgacggcgcaatgaacaccttgacggcagcagccgcccc 267
 QY 699 ggcctctgtaacctctacccggcagctgaagcagtgagcagcgtctgtagcgcg 758
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 268 ggcctctgtaacctctacccggcagctgaagcagtgagcagcgtctgtagcgcg 327
 QY 759 ccgaagcgca-cccgcgagcactcttggaacaagatctgtaagtgctctgctgcctca 817
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 328 ccgaagcgca-cccgcgagcactcttggaacaagatctgtaagtgctctgctgcctca 387
 QY 818 agtctgaagtgaatctgagccatcttgagatcttgctctgatatgtgcggaaga 877
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 388 agtctgaagtgaatctgagccatcttgagatcttgctctgatatgtgcggaaga 447
 QY 878 aaaaagaatctcggaagacttactctgacactgaactcgactcatalgaagggcgctgc 937
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 448 aaaaagaatctcggaagacttactctgacactgaactcgactcatalgaagggcgctgc 507
 QY 938 ttcgggctaigcagcccccctgcatctcaccctggccgctcctgcatctctctg 997
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 508 ttcgggctaigcagcccccctgcatctcaccctggccgctcctgcatctctctg 567
 QY 998 tgaactacccctcaac 1012
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 568 tgactacccctcac 582

RESULT 14
 AAST70846
 ID AAST70846 standard; cDNA; 981 BP.
 XX
 AC AAST70846;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #6650.
 XX

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX Homo sapiens.
 OS WO200175067-A2.
 PN 11-OCT-2001.
 PD 30-MAR-2001; 2001WO-US08631.
 PE 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 XX Drmanac RT, Liu C, Tang YT;
 P1 WPI; 2001-639362/73.
 DR P-PSDB; ABG06659.
 DR
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 1; SEQ ID No 6650; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 981 BP; 216 A; 274 C; 309 G; 182 T; 0 other;

Query Match 7.3%; Score 463.4; DB 23; Length 981;
 Best Local Similarity 97.7%; Pred. No. 3.2e-84;
 Matches 470; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1032 caagtgaagaaggtgcttaagcaagggagcactgaatgtgtgtgagccttaccatggt 1091
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 411 ccagttggaagaaggtgcttaagcaagggagcactgaatgtgtgtgagccttaccatggt 470
 QY 1092 gttgaagaaggtgagacagccaagaacaagaagactagagaagctgcgctggcgcg 1151
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 471 gttgaagaaggtgagacagccaagaacaagaagactagagaagctgcgctggcgcg 530
 QY 1152 cgagcagttctgaccccgctggcgctaccgcagctccttgctgtagaagcgctgca 1211
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 531 cgagcagttctgaccccgctggcgctaccgcagctccttgctgtagaagcgctgca 590
 QY 1212 ctggccaacatcgttagacagcgctggcgagcgtcgagcaggaactgactcggaaggcga 1271
 ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
 Db 591 ctggccaacatcgttagacagcgctggcgagcgtcgagcaggaactgactcggaaggcga 650

QY 1272 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 1331
|||||
Db 651 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 710
QY 1332 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 1391
|||||
Db 711 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 770
QY 1392 taagcagagagcgtgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 1451
|||||
Db 771 taagcagagagcgtgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 830
QY 1452 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 1511
|||||
Db 831 ggcgcgcgcacgcttgacagaccgcgcgcgttcggggccccagagaccgcgcgaagtgg 890
QY 1512 t 1512
Db 891 t 891
RESULT 15
AAI61331
ID AAI61331 standard; cDNA: 2427 BP.
AC AAI61331:
DT 22-OCT-2001 (first entry)
XX Human polynucleotide SEQ ID NO 5320.
DE Human polynucleotide SEQ ID NO 5320.
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Draeger Syndrome; chemotactic;
KW chemoketic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX Homo sapiens.
OS Homo sapiens.
XX MO200153312-A1.
XX 26-JUL-2001.
PD 26-JUL-2001.
XX 26-DEC-2000; 2000MO-US34263.
PE 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX (HVSF-) HVSF INC.
PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Dirmanac RT;
XX WPI: 2001-442253/47.
DR P-PSDB; AAM42175.
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX Claim 1; SEQ ID NO 5320; 10078bp; English.
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Draeger Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX Sequence 2427 BP; 716 A; 508 C; 524 G; 679 T; 0 other;
XX S0
Query Match 7.0%; Score 448; DB 22; Length 2427;
Best Local Similarity 67.0%; Pred. No. 5.2e-81;
Matches 653; Conservative 0; Mismatches 315; Indels 6; Gaps 1;
QY 5114 agcgcgtgaatgagctctacacagaacctacccatccttggaagccaccgtgactaca 5173
|||
Db 1 agacagttaatgagctctacacagaacctggtlccatcccatcctgaagcgcacgcgaattcc 60
QY 5174 agaagctggccgcggtgacgcgcgaactgacgagagccttcacacaaatcatgaccaga 5233
|||||
Db 61 ggaagctgacacactacactacacagctgacagagagccttcgaacagatcgtaacaag- 119
QY 5234 gtccgcgtggagagcgcgtgctcggaagattccgcgtgagccttcgaagcgcaccac 5293
|||
Db 120 -----gatcataagaagaatgtttggaacctactccgaagtgttcttggatccaaat 174
QY 5294 tcgtgacactgagatgagagagattgtgtacaagagacatcgatcacgaagctggcag 5353
|||
Db 175 ttggagatttgatgaacaagaattgtgtacaagagcctgcacattacaagcttctgt 234
QY 5354 agatctcacacgcgctgagagattctacacagagagattggagcagctgtgtaga 5413
|||||
Db 235 agatctcacacagactagagactttatgtgtacatgtttgtgacgaattgtgtgaag 294
QY 5414 ttatcaagactcttaccctgtgacacagttccaaagctgtgactcacaagaagcctacac 5473
|||
Db 295 tgattaaagactccactcctgtgtgacacaaacaaagttgacatcacaagaagcctacac 354
QY 5474 agatcagatgtgagacgcttactgtatcctcagagctaaagaagcgggtgacctact 5533
|||||
Db 355 agatcactttgtgagagcctactgtatgagatgaagaagaaggttcacatact 414
QY 5534 ttgacgcgaactatggtctgcacattcgttctgtgacgcgttcacgcgcgagtggc 5593
|||
Db 415 ttgagaagaattcaactcctcggaggttcagtacacacacccgttccactccttgaggggc 474
QY 5594 ggcgcacagggagcgtgcccagacacacaaagcgtlaagaagcgtgctcgaacgcagcaag 5653
|||
Db 475 ggcctcggggagagcgtgatgacagatgacagaagaacacagctcctgacactatgacg 534
QY 5654 ccttccttaccatcaagactcgcatcgtgtgtgcacccgggaagagacggtgtgagc 5713
|||||
Db 535 ccttccttaccatcaagacccagacagcagcttcacacgaagagagatgtttttgacac 594
QY 5714 cagtgaagtgccacatgagacatgacagagaacacagcagcagcgttccttgacacgc 5773
|||
Db 595 cgattgaagttgaccttgaagaacatgaagaagaagacccttcagttgacgttgcatat 654
QY 5774 agcagagccacacagatgataagatgataagatgtgtccttaagcgtctgttaggcaca 5833
|||||
Db 655 accagagagcgcgtgatgacaagatgattcagatgtgtcgaagcgtctgtgagacta 714
QY 5834 ccgtgaaccagagctccttgagagtgagccagaggtttttgacagagatcccggaagacc 5893
|||
Db 715 ctgtaatacagagacactggaagtagcccaagtgcttttgctgaaatttcctgcgtgac 774
QY 5894 ccaagctcttcgagatcacacaacaatgctgctctgctcaagagacttcgcagaagaat 5953

```

Db 775 caaaactctatcgacatcacacaagtgagggtatgctttaaggaattcatcatgagat 834
QY 5954 gtgaggatgctgcggaataaagccctgtgttgccggaaccaagaagataccacc 6013
Db 835 gtggtgaagctgtagaagaacaacagctctcatcacggaagaccagaaggaataatcacgc 894
QY 6014 gtgagctggaagcgcaactactgcccgtgcggaaggtctgtgcagccccctgttaccacgc 6073
Db 895 aggaactcaaaaagaactataacacagctaaagaagaacctcagccaatgatcgagcgga 954
QY 6074 gcctgcccagctg 6087
Db 955 aaatccagaactg 968

```

Search completed: October 5, 2002, 08:20:24
 Job time: 20081 sec

OY 3012 ccacagtgacattctgtcccttggtagacgggagcttgcctacagctcgtctcgggccca 3071
 Db 685 ccccgctctcgtctctcccccgtcccgctcccccgtccctcgtctcccgctctcgtctc 744
 OY 3072 ctacaaagagtgtagccacggcgctccagtcgtcccttaacagacagccctgtcagcct 3131
 Db 745 ccccgctctcccgctctcgtctcccccgtctcccccgtcgtctcctcccgctctcc 804
 OY 3112 ggcgaatgaattacccggatctctgtgcagccagcagcactcgtgacccctaacctcc 3191
 Db 805 cgtctcctcgtctctcccccgtctcccccgtcctcgtctcctcccccgtctc -ccgtc 862
 OY 3192 ctgtctgcaccccttgacactccagcctcagcctccctctgtgtccctcaacacctcca 3251
 Db 863 ctgctctcccccgtctcccccgtcctcgtctctcccccgtctcctccgctctcgtctcctccc 922
 OY 3252 gagctcacactctctccagccagcccccagccc 3284
 Db 923 gttctcccggtctcgtctcctccctcccgctctctc 955

```

RESULT 2
US-09-050-863-2/c
Sequence 2, Application US/09050863
Patent No. 6114111
GENERAL INFORMATION:
APPLICANT: Lao, Ying
APPLICANT: Hiang, Betty
APPLICANT: Payan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
TITLE OF INVENTION: System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flehm, Hobbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/050,863
FILING DATE: 30-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
IS-09-050-863-2

```

Query Match	1.0%;	Score 65.6;	DB 3;	Length 2580;
Best Local Similarity	47.0%;	Pred. No. 2.7e-05;		
Matches 235;	Conservative 0;	Mismatches 264;	Indels 1;	Gaps 1;

DY 2778 cgtacgagagccatccttcacgacgccttgttcttcttcacgcltagtgaagtat 28337
| | | | | | | | | | | | | | | |
Db 1381 CCTCTTGACCCCGGCTTCACACTCTGTCTGCCCCCTCTGCTCCTGCCCTCTTCTCT 13222

[illegible]

```

RESULT      3
US-09-359-081-2/C
Sequence 2, Application US/09359081
Patent No. 6316223
GENERAL INFORMATION:
APPLICANT: Lao, Ying
            Hiang, Betty
            Pavan, Don
TITLE OF INVENTION: Mammalian Protein Interaction Cloning
                        System
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: Pienr, Hohbach, Test, Albritton & Herbertt
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/359,081
FILING DATE: 22-Jul-1999
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/050,863
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-65636/DJB/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 949-8711
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2580 base pairs

```


LENGTH: 9600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match 1.0%; Score 65.6; DB 4; Length 9600;
Best Local Similarity 47.0%; Pred. No. 4.3e-05;
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;

QY 2778 cgtacgagagacatccctccagcagcgtgttcttcttcacagctcatggtgaagatcat 2837
DB 1427 CCTCCTGACCCCGGCGCCCTCCTCCTCCTGCCCCCTCCTGCCCCCTCCTCCT 1368
QY 2838 ggcgtgacgtctgtctgacagcagctagacacaccccgacagcgtctcccg 2897
DB 1367 GCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 1308
QY 2898 acgcttcctgagacacatcactgcttctgtgtctgtgtgagcctgagatcatcaccg 2957
DB 1307 CCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 1248
QY 2958 tctcacaagaagtgtgagcgtgagcagcactcaagcagcagcgtcttctctcaag 3017
DB 1247 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 1188
QY 3018 tgaccttcctgctcctgtgtgagcagcagcgttctgtctcagcctgtgagcagcactaa 3077
DB 1187 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 1128
QY 3078 gcaagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3137
DB 1127 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 1068
QY 3138 ggaattacccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3196
DB 1067 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 1008
QY 3197 gcccctgtacacctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3256
DB 1007 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 948
QY 3257 ccacctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3276
DB 947 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 928

RESULT 6

US-07-884-811-15/c
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: palin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Dreger, Ginger R.
;; REGISTRATION NUMBER: 33,055
;; REFERENCE/DOCKET NUMBER: 755.1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415/225-3216
;; TELEFAX: 415/952-9881
;; TELEX: 910/371-7168
;; INFORMATION FOR SEQ ID NO: 15:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 10596 bases
;; TYPE: NUCLEIC ACID
;; STRANDEDNESS: single
;; TOPOLOGY: linear
US-07-884-811-15

Query Match 1.0%; Score 65.6; DB 1; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4.4e-05;
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;

QY 2778 cgtacgagagacatccctccagcagcgtgttcttcttcacagctcatggtgaagatcat 2837
DB 2961 CCTCCTGACCCCGGCGCCCTCCTCCTCCTGCCCCCTCCTGCCCCCTCCTCCT 2902
QY 2838 ggcgtgacgtctgtctgacagcagcagcagcagcagcagcagcagcagcagcagcagc 2897
DB 2901 GCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 2842
QY 2898 acgcttcctgagacacatcactgcttctgtgtctgtgtgagcctgagatcatcaccg 2957
DB 2841 CCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 2782
QY 2958 tctcacaagaagtgtgagcgtgagcagcagcagcagcagcagcagcagcagcagcagc 3017
DB 2781 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 2722
QY 3018 tgaccttcctgctcctgtgtgagcagcagcagcagcagcagcagcagcagcagcagcagc 3077
DB 2721 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 2662
QY 3078 gcaagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3137
DB 2661 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 2602
QY 3138 ggaattacccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3196
DB 2601 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 2542
QY 3197 gcccctgtacacctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3256
DB 2541 GCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCT 2482
QY 3257 ccacctctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3276
DB 2481 CCTCCTGCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCTGCCCCCTCCT 2462

RESULT 7

US-07-885-971-15/c
; Sequence 15, Application US/07885971
; Patent No. 5328837
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco


```
Db 2661 CCGTCCTGCGCCCTGCGCCCTGCTGCGCCCTGCTGCGCCCTGCTGCTGCT 2602
Qy 3138 ggaattacacgcgacatctgtgacgacagcactagtgaccctcaacct-ccccctgt 3196
Db 2601 GCCCTCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2542
Qy 3197 gcccctgtcaccttcagcctgcgcctccctctgtgtcctcacacacaccccaagct 3256
Db 2541 GCCCTCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2482
Qy 3257 ccacctctccagccaagcc 3276
Db 2481 CCGTCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2462

RESULT 9
US-08-194-088B-15/c
; Sequence 15, Application US/08194088B
; Patent No. 5580963
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,088B
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/884811
; FILING DATE: 18-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallegos, R. Thomas
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: 755D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-2614
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-194-088B-15

Query Match 1.0%; Score 65.6; DB 1; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4.4e-05;
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;
```

```
Db 2841 CCGTCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2782
Qy 2958 tttccacaaggatgtgtagtctggccgagcaactcaagcagactgtcttctccag 3017
Db 2781 CCGTCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2722
Qy 3018 tgacctgtcccttggtgagcggggcttgtcttcagcctgtgtccggcccaactaa 3077
Db 2721 GCCCTCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2662
Qy 3078 gcaagtgagcagcgagctcagctgtccctaatccagcagccctgtgtgacctgagcat 3137
Db 2661 CCGTCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2602
Qy 3138 ggaattacacgcgacatctgtgacgacagcactagtgaccctcaacct-ccccctgt 3196
Db 2601 GCCCTCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2542
Qy 3197 gcccctgtcaccttcagcctgcgcctccctctgtgtcctcacacaccccaagct 3256
Db 2541 GCCCTCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2482
Qy 3257 ccacctctccagccaagcc 3276
Db 2481 CCGTCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCGCCCTGCT 2462
```

```
RESULT 10
US-08-194-087-15/c
; Sequence 15, Application US/08194087
; Patent No. 5679910
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,087
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: .
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/952-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-194-087-15
```

```
Query Match 1.0%; Score 65.6; DB 2; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4.4e-05;
```

Matches	235; Conservative	0;	Mismatches	264; Indels	1; Gaps	1;
Qy	2778	cgtagcagagagccatctctccagcaagcctggtcttcttccagctcatggtgaagat	2837			
Db	2961	ccctccctcaccgccgcccacacccctccctccctccctccctccctccctccctccct	2902			
Qy	2838	ggagctgacacctgctgcttggccagcgactagacacaccccgaaagctgctccccc	2897			
Db	2901	gctctgccccctctccgccccctccctccctccctccctccctccctccctccctccct	2842			
Qy	2898	acgcttccctgacacacatcactgcttggtagcttggccttggagcctgagatcaccg	2957			
Db	2841	ccctgccccctccctccctccctccctccctccctccctccctccctccctccctccct	2782			
Qy	2958	tgccacaagaatgtgagctgagccgagacactaaagccagcctgcttcttcctag	3017			
Db	2781	ccctccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2722			
Qy	3018	tgaccttctgccccctggagccggaggttggcttcaagcctgagccgagccactaa	3077			
Db	2721	gccccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2662			
Qy	3078	gaagctgacacagccgagctccagctgctccctaatccagacagccctgacccctgacat	3137			
Db	2661	ccctccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2602			
Qy	3138	ggaattcaccgacatcctgtgacgacaggaactagtgaaacctaaact-ccccgtct	3196			
Db	2601	gccccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2542			
Qy	3137	gccccctgttaacatcactcctccctccctccctccctccctccctccctccctccct	3256			
Db	2541	gccccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2482			
Qy	3257	ccaccttccacagccaagcc	3276			
Db	2481	ccctccctccctccctccctccctccctccctccctccctccctccctccctccctccct	2462			
RESULT 11						
PCT-US93-04648-15/c						
Sequence 15, Application PC/TUS9304648						
GENERAL INFORMATION:						
APPLICANT: Genentech, Inc., Godwskt, Paul J., Lokker, Natalie A., Mark, Melanie						
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS						
NUMBER OF SEQUENCES: 21						
CORRESPONDENCE ADDRESS:						
ADDRESSEE: Genentech, Inc.						
STREET: 460 Point San Bruno Blvd						
CITY: South San Francisco						
STATE: California						
COUNTRY: USA						
ZIP: 94080						
COMPUTER READABLE FORM:						
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk						
COMPUTER: IBM PC compatible						
OPERATING SYSTEM: PC-DOS/MS-DOS						
SOFTWARE: patin (Genentech)						
CURRENT APPLICATION DATA:						
APPLICATION NUMBER: PCT/US93/04648						
FILING DATE: 19930517						
CLASSIFICATION:						
PRIOR APPLICATION DATA:						
APPLICATION NUMBER: 07/884811						
FILING DATE: 13-MAY-92						
PRIOR APPLICATION DATA:						
APPLICATION NUMBER: 07/885971						
FILING DATE: 13-MAY-92						
ATTORNEY/AGENT INFORMATION:						
NAME: Dreger, Ginger R.						
REGISTRATION NUMBER: 33, 055						
REFERENCE/DOCKET NUMBER: 755,779P1						
TELECOMMUNICATION INFORMATION:						

```

: TELEPHONE: 415/225-3216
: TELEFAX: 415/952-9881
: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: NUCLEIC ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: PCT-US93-04648-15

Query Match      1.0%; Score 65.6; DB 5; Length 10596;
Best Local Similarity 47.0%; Pred. No. 4,4e-05;
Matches 235; Conservative 0; Mismatches 264; Indels 1; Gaps 1;

QY 2778 cgtacgagagagcatctccagcaagcctggtctctctccagctcatggtgaaagat 2837
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2961 CCTCGTCGACCCCGGCGCTCCACCTCCTGCTCTGCGCCCGCTGCTCCCTCCTCT 2902
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2838 ggcgcgtgaacctgctgtgtggccagcgactatagacacaccccgaaatcgcttcccg 2897
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2901 GCTCGTCGCCCCCTCCGCCCCCTCCTGCTCTGCCCCCTCGCCCCCTCCTCGGCCCT 2842
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2898 agcgttcctggagcgacatcactgacctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2957
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2841 CCTGCCCTCTCTGCTCTCTGCCCCCTCTGCCCCCTCTCTCTGCTCTGCCCCCTCT 2782
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2958 tgtccacaagaatgtgagctgtgacgagacatcaagcgacgactgttcttctcaag 3017
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2781 CCTCGCTCGCTCCGCCCCCTGCTCCCTCCTGCTCTGCTCGCCCCCTCGCCCCCT 2722
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 3018 tgaacctctgtccctgtgtgagcggagcttgtctcaagcctgtgtccgagccactaaga 3077
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2721 GCCCTCTCGCCCCCTCTGCTCTGCCCCCTCTGCTCTGCCCCCTCTGCTCTGCCCCCT 2662
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 3078 gaaagtgtgcaagcgagctcagtcgtccctaatccagcagcagctgtgaaccttgagat 3137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2661 CCGCTCTCGCCCCCTCTGCTCTGCCCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2602
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 3138 ggaattcacgcgcatctgtgtgacgacagacactagtgaaacctaaact-ccctctgt 3196
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2601 GCCCTCTCGCCCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2542
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 3197 gccccctgtcaacctcagcctcgccctccctctgtgtctctccaaacctccagagat 3256
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2541 GCCCTCTCGCCCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2482
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 3257 ccaacttctccagcaagcc 3276
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2481 CCGCTCTCGCCCCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2462
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```



```
RESULT 14
US-09-046-992-1
; Sequence 1, Application US/09046992
; Patent No. 6140066
; GENERAL INFORMATION:
; APPLICANT: Loriboum-Galski, Haya
; APPLICANT: Yarkoni, Shai
; APPLICANT: Ben-Yehudah, Ahmi
; TITLE OF INVENTION: METHODS OF CANCER DIAGNOSIS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds, LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-281.
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/046,992
; FILING DATE: 24-Mar-1998
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 9457-0013-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1908 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..1905
; OTHER INFORMATION:
US-09-046-992-1

Query Match          0.8%; Score 52.8; DB 3; Length 1908;
Best Local Similarity 45.8%; Pred. No. 0.016;
Matches 183; Conservative 0; Mismatches 217; Indels 0; Gaps 0;
```

```
Db 1205 CGGGCCCGGCGACAGCGCGCCCTGCTGAGGCGAAGTATCCCACTGGGCGGAGT 1264
Qy 5006 tgcgccgacgagagggtctctgcgcgggaagcacttcactgtggtgtgtag 5065
Db 1265 TCTCGGCGGACGCGCGGACGTCTCAGTTCAGCACCCCGCGACGACAGAACTGGACGGTGG 1324
Qy 5066 gttgtcgtgaacagcgccggtacttcaacatggtcgg 5105
Db 1325 AGCGGCTGCTCCAGGCGCACCGCACTGGAGAGCGCGG 1364

RESULT 15
US-08-356-786-15
; Sequence 15, Application US/08356786
; Patent No. 5877305
; GENERAL INFORMATION:
; APPLICANT: Huston, James S.
; APPLICANT: Oppermann, Hermann
; APPLICANT: Huston, L. L.
; APPLICANT: Ring, David B.
; TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibault
; STREET: Exchange Place, 53 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/356,786
; FILING DATE:
; CLASSIFICATION: 424
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 07/831,967
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Pitcher, Edmund R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1869
; OTHER INFORMATION: /note="product = "741sfv-DE40"
US-08-356-786-15

Query Match          0.8%; Score 51.2; DB 2; Length 1869;
Best Local Similarity 45.5%; Pred. No. 0.036;
Matches 182; Conservative 0; Mismatches 218; Indels 0; Gaps 0;
```


[illegible]

Search completed: October 5, 2002, 07:37:58
Job time: 18460 sec